格致方法

定量研究系列

· 革新研究理念

· 丰富研究工具

· 最权威、最前沿的定量研究方法指南

吴晓刚 / 主编

【美】威廉·D. 贝里(William D. Berry) / 等著

# 因果关系模型

格致出版社 上海人民出版社

世纪出版

格致方法

定量研究系列

# 因果关系模型

本书由五种讨论因果关系模型的小册子组成，分别是《非递归因果模型》、《用面板数据做因果分析》、《中介作用分析》、《多层次模型》以及《潜变量增长曲线模型》。本书主要介绍社会学研究方法之一，即因果关系模型。因果关系模型是社会科学研究中较常用的分析模型，该书通过介绍非递归因果模型、适用于因果分析的面板数据，对数据分析中常见的变量的中介作用进行阐述，对研究中普遍出现的多层次效应的使用进行讨论，以弥补普通最小二乘法在研究应用中的不足。作者还向研究者提供了一个与多层次增长模型相关的、社会科学量化应用的新方法——潜变量增长曲线模型，为研究者提供了深入了解和学习因果关系模型的途径。

上架建议：社科研究方法

ISBN 978-7-5432-1906-9

9 787543 219069 >

定价：68.00元

易文网：www.ewen.cc

格致网：www.hibooks.cn

格致方法
定量研究系列

吴晓刚 / 主编

【美】威廉·D. 贝里(William D. Berry) / 等著

# 因果关系模型

SAGE Publications ,Inc.

格致出版社 上海人民出版社

**图书在版编目(CIP)数据**

因果关系模型/(美)贝里(Berry, W. D.)等著;
吴晓刚主编. —上海:格致出版社:上海人民出版社,
2011
(格致方法·定量研究系列)
ISBN 978-7-5432-1906-9

Ⅰ. ①因… Ⅱ. ①贝… ②吴… Ⅲ. ①因果论-研究
Ⅳ. ①B025.5

中国版本图书馆 CIP 数据核字(2010)第 262174 号

责任编辑 顾 悦
封面装帧 人马艺术工作室·储平

---

**因果关系模型**
[美]威廉·D.贝里 等著
吴晓刚 主编

---

| | | |
|---|---|---|
| **出 版** | 世纪出版集团 www.ewen.cc | 格致出版社 www.hibooks.cn 上海人民出版社 |
| | (200001 上海福建中路193号24层) | |
| | 编辑部热线 021-63914988<br>市场部热线 021-63914081 | |
| **发 行** | 世纪出版集团发行中心 | |
| **印 刷** | 上海江杨印刷厂 | |
| **开 本** | 787×1092 毫米 1/16 | |
| **印 张** | 28.5 | |
| **插 页** | 1 | |
| **字 数** | 412,000 | |
| **版 次** | 2011 年 7 月第 1 版 | |
| **印 次** | 2011 年 7 月第 1 次印刷 | |

ISBN 978-7-5432-1906-9/C·50

**定 价** 68.00元

# SAGE 绿皮书的传奇

1976 年，当政府管理与政治学教授 Eric Uslaner 向 SAGE 出版社的创始人 Sara 建议出版关于重要定量研究方法的简明小册子时，没有人预见到这套书会取得巨大成功。

那年夏天，"社会科学定量研究方法" 丛书诞生。统一的、朴素的绿色封面，上面仅仅印着书名、作者名及本书所属的系列名。Iverson 和 Northporth 的《变量分析》( *Analysis of Variance* )、Nagel 的《运筹学》( *Operations Research* ) 以及 Henkel 的《显著性检验》( *Tests of Significance* ) 是最早出版的几本，每本售价 2.95 英镑，被形容为 "像烤饼一样好卖"。

SAGE 选择了 20 种顶尖的定量研究工具，然后去寻找合适的作者，邀请他们围绕这些工具撰写 92 页的小书。这些薄薄的绿皮书在课堂上深受欢迎，在图书馆成套陈列。

至今，绿皮书系列中共有 160 种在初版或再版。书的主题反映了量化研究方法的发展：从基础统计知识、数据类型、测量到计算机的应用以及博弈论。这套书非常畅销，其中最畅销的一本，是 1980 年出版的 Michael Lewis-Beck 的《回归方法的应用》( *Applied Regression* )。

# 出版说明

本书由五种讨论因果关系模型的小册子组成，分别是《非递归因果模型》、《用面板数据做因果分析》、《中介作用分析》、《多层次模型》以及《潜变量增长曲线模型》。本书的主要内容如书名所示，是介绍社会学研究方法之一——因果关系模型。因果关系模型是社会科学研究中较复杂的分析模型，该书通过介绍非递归因果模型、适用于因果分析的面板数据，对数据分析中常见的变量的中介作用进行阐述，并对研究中普遍出现的多层次效应的使用进行讨论，以弥补普通最小二乘法在研究应用中的不足。作者还向研究者提供了一个与多层次增长模型相关的、社会科学量化应用的新方法——潜变量增长曲线模型，为研究者提供了深入了解和学习因果关系模型的新途径。

《非递归因果模型》针对因果效应往往并非单向这一问题，提出非递归模型，并部分涉及辨识方程系统的问题；《用面板数据做因果分析》通过运用大量真实的案例，阐明了面板数据带来的因果推论机会，并展示了克服各种分析障碍的方法；《中介作用分析》指出，中间变量的中介作用在理解个体和社会方面具有重要作用，并运用线性回归、有向非循环图和结构方程模型三种方法来考察因果关系；《多层次模型》针对个体存在于群体中时，不相关的误差项假设无法成立的情况，提出多层次模型作为解决方法；《潜变量增长曲线模型》亦是在横截面数据和纵向分析模型存在不足的情况下，提出以潜变量增长曲线模型作为对策，说明其表达和估计的方式，并讨论了该模型的延伸，如分段增长、结构潜在增长等。

# 总序

往事如烟，光阴如梭。转眼间，出国已然十年有余。1996 年赴美留学，最初选择的主攻方向是比较历史社会学，研究的兴趣是中国的制度变迁问题。以我以前在国内所受的学术训练，基本是看不上定量研究的。一方面，我们倾向于研究大问题，不喜欢纠缠于细枝末节。国内一位老师的话给我的印象很深，大致是说：如果你看到一堵墙就要倒了，还用得着纠缠于那堵墙的倾斜角度究竟是几度吗？所以，很多研究都是大而化之，只要说得通即可。另一方面，国内（十年前）的统计教学，总的来说与社会研究中的实际问题是相脱节的。结果是，很多原先对定量研究感兴趣的学生在学完统计之后，依旧无从下手，逐渐失去了对定量研究的兴趣。

我所就读的美国加州大学洛杉矶分校社会学系，在定量研究方面有着系统的博士训练课程。不论研究兴趣是定量还是定性的，所有的研究生第一年的头两个学期必须修两门中级统计课，最后一个学期的系列课程则是简单介绍线性回归以外的其他统计方法，是选修课。希望进一步学习定量研究方法的可以在第二年修读另外一个三学期的系列课程，其中头两门课叫"调查数据分析"，第三门叫"研究设计"。除此以外，还有如"定类数据分析"、"人口学方法与技术"、"事件史分析"、"多层线性模型"等专门课程供学生选修。该学校的统计系、心理系、教育系、经济系也有一批蜚声国际的学者，提供不同的、更加专业化的课程供学生选修。2001 年完成博士学业之后，我又受安德鲁·梅隆基金会资助，在世界定量社会科学研究的重镇密歇根大学从事两年的博士后研究，其间旁听谢宇教授为博士生讲授的统计课程，并参与该校社会研究院（Institute for Social Research）定量社会研究方法项目的一些讨论会，受益良多。

2003年，我赴港工作，在香港科技大学社会科学部，教授研究生的两门核心定量方法课程。香港科技大学社会科学部自创建以来，非常重视社会科学研究方法论的训练。我开设的第一门课"社会科学里的统计学"(Statistics for Social Science)为所有研究型硕士生和博士生的必修课，而第二门课"社会科学中的定量分析"为博士生的必修课(事实上，大部分硕士生在修完第一门课后都会继续选修第二门课)。我在讲授这两门课的时候，根据社会科学研究生的数理基础比较薄弱的特点，尽量避免复杂的数学公式推导，而用具体的例子，结合语言和图形，帮助学生理解统计的基本概念和模型。课程的重点放在如何应用定量分析模型研究社会实际问题上，即社会研究者主要为定量统计方法的"消费者"而非"生产者"。作为"消费者"，学完这些课程后，我们一方面能够读懂、欣赏和评价别人在同行评议的刊物上发表的定量研究的文章；另一方面，也能在自己的研究中运用这些成熟的方法论技术。

上述两门课的内容，尽管在线性回归模型的内容上有少量重复，但各有侧重。"社会科学里的统计学"(Statistics for Social Science)从介绍最基本的社会研究方法论和统计学原理开始，到多元线性回归模型结束，内容涵盖了描述性统计的基本方法、统计推论的原理、假设检验、列联表分析、方差和协方差分析、简单线性回归模型、多元线性回归模型，以及线性回归模型的假设和模型诊断。"社会科学中的定量分析"则介绍在经典线性回归模型的假设不成立的情况下的一些模型和方法，将重点放在因变量为定类数据的分析模型上，包括两分类的 logistic 回归模型、多分类 logistic 回归模型、定序 logistic 回归模型、条件 logistic 回归模型、多维列联表的对数线性和对数乘积模型、有关删节数据的模型、纵贯数据的分析模型，包括追踪研究和事件史的分析方法。这些模型在社会科学研究中有着更加广泛的应用。

修读过这些课程的香港科技大学的研究生，一直鼓励和支持我将两门课的讲稿结集出版，并帮助我将原来的英文课程讲稿译成了中文。但是，由于种种原因，这两本书拖了四年多还没有完成。世界著名的出版社 SAGE 的"定量社会科学研究"丛书闻名遐迩，每本书都写得通俗易懂。中山大学马骏教授向格致出版社何元龙社长推荐了这套书，当格致出版社向我提出从这套丛书中精选

一批翻译，以飨中文读者时，我非常支持这个想法，因为这从某种程度上弥补了我的教科书未能出版的遗憾。

翻译是一件吃力不讨好的事。不但要有对中英文两种语言的精准把握能力，还要有对实质内容有较深的理解能力，而这套丛书涵盖的又恰恰是社会科学中技术性非常强的内容，只有语言能力是远远不能胜任的。在短短的一年时间里，我们组织了来自中国内地及港台地区的二十几位研究生参与了这项工程，他们目前大部分是香港科技大学的硕士和博士研究生，受过严格的社会科学统计方法的训练，也有来自美国等地对定量研究感兴趣的博士研究生。他们是：

香港科技大学社会科学部博士研究生蒋勤、李骏、盛智明、叶华、张卓妮、郑冰岛，硕士研究生贺光烨、李兰、林毓玲、肖东亮、辛济云、於嘉、余珊珊，应用社会经济研究中心研究员李俊秀；香港大学教育学院博士研究生洪岩璧；北京大学社会学系博士研究生李丁、赵亮员；中国人民大学人口学系讲师巫锡炜；中国台湾“中央”研究院社会学所助理研究员林宗弘；南京师范大学心理学系副教授陈陈；美国北卡罗来纳大学教堂山分校社会学系博士候选人姜念涛；美国加州大学洛杉矶分校社会学系博士研究生宋曦。

关于每一位译者的学术背景，书中相关部分都有简单的介绍。尽管每本书因本身内容和译者的行文风格有所差异，校对也未免挂一漏万，术语的标准译法方面还有很大的改进空间，但所有的参与者都做了最大的努力，在繁忙的学习和研究之余，在不到一年的时间内，完成了三十五本书、超过百万字的翻译任务。李骏、叶华、张卓妮、贺光烨、宋曦、於嘉、郑冰岛和林宗弘除了承担自己的翻译任务之外，还在初稿校对方面付出了大量的劳动。香港科技大学霍英东南沙研究院的工作人员曾东林，协助我通读了全稿，在此我也致以诚挚的谢意。有些作者，如香港科技大学黄善国教授、美国约翰·霍普金斯大学郝令昕教授，也参与了审校工作。

由于所选每本书都有一篇序言，对相关方法的背景和应用作了很好的介绍，我们均予以保留，内容在此不再赘述。为了方便起见，我们将内容相似的书目集册出版，每册三至五本不等，共八册，它们分别是：《线性回归分析基础》、

《高级回归分析》、《广义线性模型》、《列表数据分析》、《纵贯数据分析》、《因果关系模型》、《社会科学中的数理基础及应用》和《数据分析方法五种》。所冠书名未必能精准涵盖其中的内容，读者可自行参阅每本书的序言或目录。

我们希望本丛书的出版，能为推动国内社会科学定量研究的扎实学风作出一点贡献。

吴晓刚

于香港九龙清水湾

# 目录

## 非递归因果模型


序 3

第 1 章　导论 4

第 2 章　可辨识性问题 14

第 3 章　可辨识性的检验 32

第 4 章　修改不可辨识模型 47

第 5 章　估计方法 51

第 6 章　结论 68

附录 71

注释 77

参考文献 80

译名对照表 82


## 用面板数据做因果分析


序 87

第 1 章　导论 89

第 2 章　用面板数据对变化进行建模 92

第 3 章　交互因果模型 110

第 4 章　测量误差模型 130

第 5 章　虚假相关及自相关的扰动项 152
第 6 章　关于在面板分析中进行因果推论的结语 167
附录 168
注释 170
参考文献 172
译名对照表 176

## 中介作用分析

序 181
第 1 章　中介作用简介 182
第 2 章　中介作用分析的基本方法 192
第 3 章　用结构方程模型来分析中介作用:测量模型 197
第 4 章　用结构方程模型来分析中介作用:结构模型 213
第 5 章　高级议题 230
第 6 章　结论 249
附录 251
注释 261
参考文献 266
译名对照表 271

## 多层次模型


序 275

第 1 章 为什么使用多层次模型 276

第 2 章 基本多层次模型 282

第 3 章 基本多层次模型的扩展 317

附录 333

参考文献 336

译名对照表 339


## 潜变量增长曲线模型


序 343

第 1 章 简介 345

第 2 章 将潜变量增长曲线模型应用于经验数据 365

第 3 章 模型的特别拓展 398

第 4 章 潜变量增长模型和多层模型的关系 412

第 5 章 结论 421

附录 424

注释 426

参考文献 428

译名对照表 440

# 非递归因果模型

## 作 者 简 介

### 威廉·D. 贝里(William D. Berry)

美国佛罗里达州立大学政治科学系教授。其研究领域是公共政策、政治经济学和国家政治学，并是《多元回归的实际应用》(*Multiple Regression in Practice*)和《理解美国政府的膨胀》(*Understanding United States Government Growth*)的合著者，还就研究方法在《美国政治科学评论》(*American Political Science Review*)、《美国政治学杂志》(*American Journal of Political Science*)和《政治学杂志》(*Journal of Politics*)等期刊上发表了大量论文。

## 译 者 简 介

### 洪岩璧

曾先后获得中国人民大学社会学系法学学士学位及北京大学社会学硕士学位，现为香港大学教育学院博士候选人。主要研究领域包括社会分层与教育不平等、农村与少数族群的教育发展。

### 陈陈

香港大学教育学院博士毕业，现为南京师范大学心理学系副教授。

# 序

随着社会科学理论的日益复杂化，我们亟须反映此种复杂性的统计模型来补充先前最为简单的模型。这一趋势适用于回归模型，因为在回归模型中，研究已经逐渐从单方程模型转向多方程模型。有时候，当因果效应被认为是单向的，并允许使用所谓的递归模型时，多方程模型自身也就相对简单。但大多数时候，单向因果作用的假设是不现实的，因此，研究者转向了非递归模型。

随着上述发展的出现，全面且深入浅出地介绍和讨论这些更复杂的技术的需求就显得尤为迫切。更重要的是，这类著作应当包含对下列过程的一种真正的理解：模型的使用、模型所依据的假设、模型结果的解释等等。威廉·D. 贝里的《非递归因果模型》无疑正是这样的佳作。

贝里假定读者基本熟悉递归模型，比如已经读过早先出版的赫伯特·阿什（Herbert Asher）的《因果模型》（*Causal Modeling*）一书。为了唤起读者的记忆，同时搭建对本书内容理解的合适桥梁，贝里先在第 1 章中简要回顾了递归模型的基本假设，然后才转入讨论非递归模型如何被应用于估计更为复杂的方程系统。

本书的其余部分大都聚焦于“辨识”问题。因为辨识方程系统是研究工作中最困难的部分，故而努力把这一过程表述清楚成为本书重要的组成部分。这本介绍性的著作还有一个优点：贝里在解释书中材料时，并不要求读者具备线性代数知识。

在最后一章中，贝里回顾了那些已经被确认用于估计非递归模型参数的通用技术。纵观全书，作者应用了来自经济学、政治科学和社会学的一系列例子。

理查德·G. 尼米（Richard G. Niemi）

# 第 1 章 | 导论

社会科学经验研究中，最常用的策略是建立一个单方程模型，然后利用一个样本数据来估计方程中的系数。在这个单方程模型中，一个变量被定义为因变量，它被假定受一个或多个解释变量(自变量)的影响。比如，某人可以建立一个有关选举中投票行为的单方程模型，它认为个人在某次特定的选举中对互为对手的候选人的比较性评价是因变量，该因变量有两个解释变量：某人的政策或议题立场($X_1$)和某人的政党认同感($X_2$)，再加上一个误差项决定。[1]

$$对候选人评价 = a + b_1X_1 + b_2X_2 + e \qquad [1.1]$$

接下来，如果我们从一个样本中获得了这 3 个变量的一组数据，而且愿意为进行这个模型所必要的回归分析作出一组特定假设(参见 Lewis-Beck, 1980)，那么我们就可以用一般最小二乘法(OLS)回归分析来估计方程的系数 a、$b_1$ 和 $b_2$。这个对于解释变量的系数估计值可以被解释为当其他解释变量的系数估计值为常量的情况下，该变量对因变量(候选人评价)的直接效应。比如，假定政党认同保持不变，那么 $b_1$ 的估计值就代表了政策或观点立场对候选人评价的直接作用。

但是，很多社会科学的理论表明，变量之间的因果关系过于复杂，几乎不可能借助单方程模型来反映。[2] 比如，假定某模型中的因变量受多个解释变量影响，模型背后的理论可能认为其中一些解释变量是另一些解释变量的原因。例如在投票行为中，我们可以修改单方程模型的方程 1.1，认为个人的政策/议题立场变量很可能影响其政党认同。对这两个解释变量的预期也许基于这样一个信念：隶属于某一政党的人们对政党议题的看法和基本的意识形态立场受到

其所认同的政党领导人和候选人的引导，至少部分如此。把这个修改加到先前的模型中去，就产生了图 1.1 所示的因果关系模型。[3]这一模型认为，政党认同和政策/议题立场决定了对候选人的评价的因素，但同时政策/议题立场又直接影响个体的政党认同。就此而言，政策/议题立场不仅直接（如先前的模型方程 1.1 所示）而且间接（通过它对政党认同的作用）影响对候选人的评价。图 1.1 所示的关于候选人选择的理论无法仅仅用一个方程进行模型化，而是需要一个多方程模型。

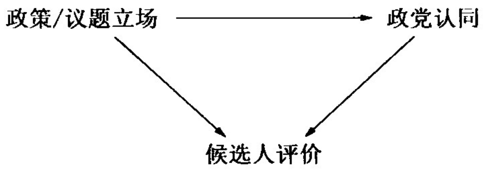



**图 1.1　一个变量之间因果关系的“因果模式”：以投票行为为例**

的确，多方程（或者说因果）模型被频繁使用于社会科学研究中，得到了丰硕的研究结果。很多社会科学家想必对多方程模型的一个子类，即常用的递归因果模型非常熟悉。在过去的 20 多年中，该模型被社会学家和政治科学家频繁使用（参见 Land, 1969; Duncan et al., 1971）。所谓递归，就是一个模型必须同时满足几个条件，以确保模型中界定的所有因果效应在本质上都是单向的，亦即模型中任何两个变量之间都不互为相关，一个变量不影响另一个变量。另外，模型中所有的误差项（或干扰项）都必须假定为互不相关。递归模型的一个优点是模型中的系数估计非常方便。所有的递归模型都是“可辨识的”。“可辨识”一词的确切含义我们要到下一章才会讨论。现在可以指出的是，为了能够确切估计多方程模型中的系数，从而为因果效应的本质提供有意义的信息，该模型首先必须是可辨识的。进一步而言，递归模型的假定允许我们使用一般最小二乘法回归分析获得模型系数的无偏估计值。[4]因此，受过回归分析训练的社会科学家可以轻松地进入递归模型分析。

但是，递归模型中的前提假设往往与我们所研究的社会科学过程的本质相矛盾。在诸多情况下，假定模型中没有互为相关（互为因果）的两个变量是不现

实的。而且，考虑到我们通常在很大程度上忽略了模型中干扰项所代表的因素，要为误差项互不相关这一假设提供令人信服的证明也常常是不可能的。鉴于这些情况，我们必须放弃递归模型，采用非递归多方程（或同步方程）模型。这类模型允许变量之间存在互为因果关系，而且假定一对或多对误差项之间存在非零相关。

非递归模型在经济学中已被广泛运用了几十年，是几乎所有计量经济学教科书的一个核心部分（如 Klein，1962；Christ，1966；Theil，1971）。但政治科学家和社会学家却很少采用此类模型，除了一些例外，如政治科学家 Hibbs（1973）、Jackson（1975）和 Erikson（1976）以及社会学家 Mason 和 Halter（1968）、Land（1971）、Waite 和 Stolzenberg（1976），他们的研究都采用了非递归模型。研究者较少使用非递归模型，是因为它的使用带来了递归模型不会产生的一些研究问题。比如，一些非递归模型是“不可辨识的”。我们将在下一章中正式定义这一概念。这里要说明的是，对于不可辨识的模型，即使有最好的数据，我们也不可能获得模型系数的有意义的估计值。

“可辨识”的意义以及导致一些非递归模型“不可辨识”的因素分析将是本书第 2 章的主题。幸运的是，存在一些可以检测模型是否“可辨识”的简便方法，我们将在第 3 章和第 4 章中介绍它们。在第 5 章中，我将讨论如何把不可辨识模型修改成可辨识模型的一些策略。在第 6 章中，我将介绍适用于可辨识的非递归模型的系数估计的步骤。虽然适用于递归因果模型的一般最小二乘法不适用于非递归模型，但幸运的是，我们可以对一般最小二乘法回归进行修改，使它能为非递归模型提供具有一致性的系数估计。[5]

对非递归同步方程模型更为详尽的讨论需要线性代数知识，因为检验某个模型是否可辨识的最好方法是秩条件，而这一方法传统上需要用线性代数来分析。然而，本书提供了一个不需要读者具备线性代数知识的运算法则来使用秩条件。虽然不要求线性代数知识，但本书要求读者具备一般最小二乘法回归分析的知识，并且大体上熟悉递归因果模型。[6]

## 概念定义和符号标记法

多方程模型的分析需要对一些重要概念进行定义，并发展一些表示这类模型的符号标记法。首先，我们必须区分多方程模型中用到的几种变量。内生变量是在模型中被清晰地表示出来的原因变量，因此它们是模型背后的理论所要解释的变量。相反，前定变量是那些没有在模型中被清晰地表示出来的原因变量。换言之，这些变量的值是由外在于模型的因素决定的，因此它们被认为是给定的，是模型背后的理论不试图解释的。[7]前定变量可以进一步被区分为两类：第一类是滞后内生变量，它的值等于模型在前一时间点的内生变量的值；第二类是外生变量，它完全由模型以外的因素决定，且并不简单等于前一时间点的内生变量值。鉴于在大部分社会科学研究中，所有的前定变量都是外生的，因此，在大多数情况下，我将只简单区分内生变量和外生变量。但是，当我使用“外生变量”一词时，我都在更精确的意义上指代更广义的前定变量。[8]

我将用 X 来标记本书中的内生变量：用 $X_1$，$X_2$，…，$X_m$ 标记模型中的一组内生变量（因为是多方程模型，所以 $m \geqslant 2$）。另外，我将用 Z 标记外生变量，以 $Z_{m+1}$，$Z_{m+2}$，…，$Z_{m+k}$ 表示模型中的 k 个外生变量（$k \geqslant 1$）。运用这种标记法，每个内生变量直接受模型中所有其他变量影响的一个同步方程模型就是通常所说的饱和非递归模型，它可以被表示为 m 个结构方程组，每个方程表示一个内生变量 $X_i$ 受模型中所有其他变量的直接因果作用[9]：

$$
\begin{aligned}
X_1 &= \beta_{12}X_2 + \cdots + \beta_{1m}X_m + \gamma_{1,m+1}Z_{m+1} + \cdots + \gamma_{1,m+k}Z_{m+k} + \varepsilon_1 \\
X_2 &= \beta_{21}X_1 + \cdots + \beta_{2m}X_m + \gamma_{2,m+1}Z_{m+1} + \cdots + \gamma_{2,m+k}Z_{m+k} + \varepsilon_2 \\
&\vdots \\
X_m &= \beta_{m1}X_1 + \beta_{m2}X_2 + \cdots + \beta_{m,m-1}X_{m-1} + \gamma_{m,m+1}Z_{m+1} + \cdots + \\
&\quad \gamma_{m,m+k}Z_{m+k} + \varepsilon_m
\end{aligned}
\qquad [1.2]
$$

在这些结构方程中，$\beta_{ij}$标记的系数或参数表示内生变量 $X_j$ 对内生变量 $X_i$ 的直接作用，即当模型中所有其他变量保持为常量时，$X_j$ 每增加一个单位，$X_i$ 的变

化量。$\gamma_{ij}$则表示外生变量 $Z_j$ 对内生变量 $X_i$ 的直接作用，即当控制所有其他变量时，$Z_j$ 每变化一个单位，$X_i$ 的变化量。最后，结构方程中定义 $X_i$ 的 $\varepsilon_i$ 是误差项（或干扰项），表示没有被包括在模型中的所有其他变量对内生变量 $X_i$ 的作用。

## 递归因果模型：简要回顾

通过对方程组 1.2 的饱和非递归模型的参数和误差项的性质进行一些非常严格的假设，我们就能从饱和非递归模型中发展出一个递归因果模型，如果这个模型满足“递归”所需的几个条件。

首先，这个模型必须是分层的。如果模型中所有的内生变量能够按照 $X_1$，$X_2$，…，$X_m$ 排列和标记，且对于任何的 $X_i$ 和 $X_j$，当 $i<j$ 时，$X_j$ 都不能被视为 $X_i$ 的原因，那么这个模型就是分层的。故而 $\beta_{ij}$必须等于 0。因此，模型中的内生变量必须具有一定的次序，第一个内生变量仅由外生变量决定，第二个内生变量仅由外生变量和第一个内生变量决定，第三个内生变量仅由外生变量和第一、第二个内生变量决定，以此类推。这样，在递归模型中，没有任何两个内生变量是互为相关、彼此直接互为因果的。其次，也不允许存在任何间接的因果联系，从而确保在因果顺序上，任何内生变量都不能影响次序上在它之前的内生变量。最后，除了模型要成为分层模型外，所有的递归模型都必须假定每个误差项与所有的外生变量、模型中所有其他误差项都不相关。用数学形式表述就是，对于 1 和 m 之间的所有 i 和 j 的值，$cov(\varepsilon_i, \varepsilon_j)$必须为 0；对于 1 和 m 之间的所有 i 值以及 $m+1$ 和 $m+k$ 之间的所有的 j 值，$cov(\varepsilon_i, Z_j)$必须为 0。

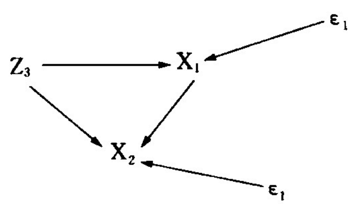
一个递归模型的所有前提假设放在一起就是：当内生变量按分层次序被标记为 $X_1$，$X_2$，…，$X_m$ 时，对于所有的 $i<j$，$cov(X_i, \varepsilon_j)=0$(Namboodiri et al.，1975：444—448；Duncan，1975：第 4 章)。因此，一个特定方程的误差项，例如 $\varepsilon_j$ 必须与内生变量 $X_1$，$X_2$，…，$X_{j-1}$不相关。但既然模型中每个误差项与模型的所有外生变量不相关（先前的假设），$\varepsilon_j$ 就会与模型中所有包含 $\varepsilon_j$ 的解释变量不相关。正是递归模型的这一特性，使研究者可以运用一般最小二乘法回归分析

来估计递归模型的系数，并获得无偏且一致的估计值。[10]因此，由于递归模型具有参数估计便利的优点，所以我们可以用处理单方程回归模型的方法来处理递归模型。但在多数情况下，递归模型的这些假定是不现实的。以图 1.1 所示的投票行为模型为例，该递归模型的方程表达式如下：

$$X_1 = \gamma_{13} Z_3 + \varepsilon_1 \quad [1.3]$$

$$X_2 = \beta_{21} X_1 + \gamma_{23} Z_3 + \varepsilon_2 \quad [1.4]$$

其中，$Z_3$ 代表个人的政策/议题立场，是模型中唯一的外生变量；两个内生变量分别是个人政党认同（$X_1$）和对候选人的评价（$X_2$）。我们的模型满足递归模型必须分层的要求，两个内生变量依次排列为“$X_1$，$X_2$”，其中 $X_1$（政党认同）仅由外生变量 $Z_3$ 决定，而 $X_2$（候选人评价）仅受 $Z_3$ 和 $X_1$ 的影响。因为模型是递归的，所以我们必须假定 $cov(\varepsilon_1, \varepsilon_2) = 0$，而且这两个误差项与外生变量 $Z_3$ 不相关。[11]这一递归模型可以用图 1.2 的因果图示来表示。




注：假设 $cov(\varepsilon_1, \varepsilon_2) = cov(\varepsilon_1, Z_3) = cov(\varepsilon_2, Z_3) = 0$。$X_1$ 为政党认同，$X_2$ 为候选人评价，$Z_3$ 为政策/议题立场。

**图 1.2　方程 1.3 和方程 1.4 所示模型的因果图示**

然而，很多人会质疑递归模型所需的这些假定在这里并不能得到证明。首先，模型的分层特性就存在疑问。举例来说，在选举中，个人的政党认同可能会影响其对候选人的评价；对候选人的评价也会修正个人对候选人所代表的政党的态度，从而影响其政党认同。我们同样也可以质疑模型的另一个假定，即对候选人的评价和政党认同都不影响政策/议题立场。在议题立场的形成上，人们可能从他们所支持的政治领导人和政党中获取线索。如果确实如此，那么我们就必须允许候选人评价和政党认同对政策/议题立场产生直接作用。如果遵从上述假设，我们的投票行为模型就不再是分层的了。事实上，将这些新增的

因果效应的假设考虑进这个递归模型后，政策/议题立场将变成内生变量，模型中每一个内生变量都将被看做所有其他内生变量的原因，如图 1.3 所示。

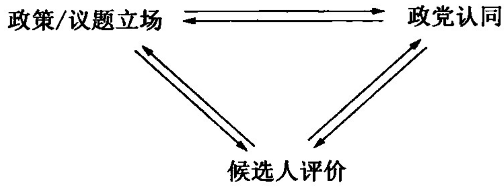



**图 1.3　表示变量间互为因果关系的箭头因果图式：以投票行为为例**

我们还可以质疑，递归模型所要求的关于误差项的假定对于图 1.2 中的模型是否合理。我之前已经提到，一个误差项可被看做代表了未被纳入模型的变量的效应。基于这一构想，我们来看 $\varepsilon_1$ 和 $\varepsilon_2$ 互不相关的假定。如果这一假定成立，那么我们就必须相信那些影响个人政党认同但未被纳入模型的因素与那些影响候选人评价，但也同样未被纳入模型的变量互不相关。同样，为了接受误差项与外生变量 $Z_3$ 互不相关的假定，我们必须相信误差项所代表的“被忽略的因素”与个人政策/议题立场互不相关。

Page 和 Jones(1979)提出了图 1.2 所示的包含 3 个变量的递归模型，用以分析总统选举中的投票行为。但 Page 和 Jones 的模型包含一个假定既影响个人(当前)的政治认同，又影响其候选人评价的变量——个人的“政党投票史”，即在先前的总统选举中，个人支持某个政党的一致性程度。如果 Page 和 Jones 是正确的，那么政党投票史就是反映在图 1.2 所示的模型中的 $X_1$(政党认同)和 $X_2$(候选人评价)的误差项中的一个因素。这样，我们可以预期 $\varepsilon_1$ 和 $\varepsilon_2$ 是相关的，因而导致了递归模型的前提假设不合理。Page 和 Jones 还提出，教育和意识形态(在左—右派的连续谱上衡量)这两个变量对政党认同和政策/议题立场都有直接作用。如果确实如此，那么递归模型关于方程 1.3 中的误差项 $\varepsilon_1$ 与外生变量 $Z_3$(政策/议题立场)不相关的假设就不成立了。

具有理论重要性变量的缺失可能导致模型中误差项彼此相关，而且存在测量误差也会导致误差项之间的相关。除非模型中变量的测量极其完善，否则测量误差将始终是误差项的组成部分。我们一般用相似的测量工具来测量模型

中的几个内生变量,那么测量工具本身带来的系统误差就会以相似的方式存在于每个变量中,这样,误差项之间的相关就不可避免。以图 1.2 的投票模型为例,如果该模型中的变量采用来自同一调查的数据,那么下述几个方面可能导致方程中误差项之间的互为相关:(1)用来测量模型中几个变量的问题,在表述上具有相似性;(2)访谈人的特性;(3)在对问题答案的编码过程中产生的相似错误;(4)其他来自调查的特点。

对于图 1.2 所示的投票行为递归模型,如果我们接受上述任何这些针对其前提假设的挑战,并因此放宽了模型的假设条件,使其更贴近现实,那么该模型将不再是递归的。在这种情况下,该模型也不能被辨识,模型参数的有意义的估计值也就无法获得。

总而言之,我们有足够的理由怀疑递归模型所需的严格假设是否合适。因此,我们不应当轻易地,或者仅仅为了便利的目的而使用递归模型,除非我们确信以下几点:(1)变量之间的因果关系的确是单向的;(2)模型中构成误差项的因素对每一个方程而言都不相同。相反,我们应当转而努力开发更贴近现实的非递归模型。当递归模型所需的假设被违背而我们仍然使用递归模型,并用一般最小二乘法去估计模型的系数时,就会得到有偏差且不一致的估计值,从而影响对因果作用大小评估的准确性。

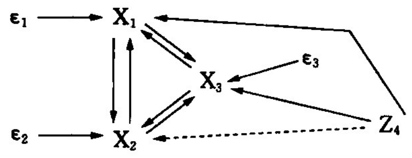
## 非递归模型的前提假设

在方程组 1.2 中,我们已经看到了包含 m 个内生变量和 k 个外生变量的饱和非递归模型。在饱和非递归模型中,每个内生变量都被假定受到模型中所有其他内生变量和外生变量的直接影响。我们可以建立一个关于投票选择的饱和非递归模型,包含图 1.3 所示的因果关系,然后加入教育水平这一外生变量对模型中内生变量的作用。为谨慎起见,我们假定教育($Z_4$)对模型中所有 3 个内生变量都具有直接作用。这样就产生了如图 1.4 的箭头图所描述的模型(在这里,虚线箭头与实线箭头并无区别),它可用下列方程表示:

$$X_1 = \beta_{12}X_2 + \beta_{13}X_3 + \gamma_{14}Z_4 + \varepsilon_1 \quad [1.5]$$

$$X_2 = \beta_{21}X_1 + \beta_{23}X_3 + \gamma_{24}Z_4 + \varepsilon_2 \quad [1.6]$$

$$X_3 = \beta_{31}X_1 + \beta_{32}X_2 + \gamma_{34}Z_4 + \varepsilon_3 \quad [1.7]$$




注：假设 $cov(\varepsilon_1, Z_4) = cov(\varepsilon_2, Z_4) = cov(\varepsilon_3, Z_4) = 0$。$X_1$ 为政党认同，$X_2$ 为候选人评价，$X_3$ 为政策/议题立场，$Z_4$ 为教育。

**图 1.4 方程 1.5、方程 1.6 和方程 1.7 所示的完全非递归模型的因果图示**

但有效应用于经验研究中的非递归模型不能是饱和的，因为饱和的非递归模型(正如下面我们要看到的)是不可辨识的。因此，我们通常假定非递归模型中的某些参数等于 0，也就是说，模型中的某些变量被假定为对一个或多个内生变量没有直接的因果作用。

举例来说，Page 和 Jones(1976)假定，教育对候选人的评价没有直接作用，而且教育对候选人评价的所有影响都是间接通过教育对政策/议题立场和政党认同的影响来实现的。如果我们接受这一个假定，我们就可以建立一个新的非饱和模型，其中 $\gamma_{24} = 0$。只要把图 1.4 中的虚线箭头删除，该箭头图就可用来表示我们现在所讨论的模型。它的方程组形式是保留方程 1.5 和方程 1.7，但以下面的方程替代方程 1.6：

$$X_2 = \beta_{21}X_1 + \beta_{23}X_3 + \varepsilon_2 \quad [1.8]$$

细心的读者也许会发现，我们至今尚未对方程组 1.2 所表示的模型或刚刚讨论的关于投票选择的非递归模型中的误差项作出任何假设。因此，我们有必要提出一些清晰的假设。这些假设的特性将极大地影响一个模型是否可辨识。在本书中(除了附录 2)，我们将把对非递归模型的讨论限定在下述典型的假设中：(1)对于所有的 i 和 j 来说，$cov(\varepsilon_i, Z_j) = 0$ (即模型中的每一个误差项与所有的外生变量都不相关)；(2)对于所有的 i，$E(X_i) = E(Z_j) = E(\varepsilon_i) = 0$ (即所有

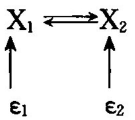
变量和误差项的均值都为 0)。假设(2)的设定是为了表述的方便。设定变量的测量尺度的原点只是一种惯例,便于我们消除模型方程中的常数项。

一方面,假设(1)对于一个非递归模型是否可辨识具有重要作用。尽管该假设在很多情况下是合理的,但如果研究者在某个特定模型中不作出这一假设,那么本书所讨论的模型可辨识性的检验方法将无用武之地。另一方面,在某些情况下,研究者也许会作出比假设(1)更强的有关误差项的假定。例如,除了假定每个误差项与模型中所有外生变量不相关之外,研究者还可以假定某些误差项之间互不相关。在这种情况下,本书讨论的模型可辨识与否的检验方法依然不适用(除非进行适当修改)。有一种称为“组群递归模型”的非递归模型,它仅假定某些误差项之间互不相关。有关讨论请参阅附录 2。

但是我们要切记,在非递归模型中假定每个误差项与所有内生变量都不相关是没有意义的,因为这类模型中至少会有一个误差项与一个或多个内生变量相关。以图 1.4 模型中的方程 1.5、方程 1.6 和方程 1.7 为例,假定 $\varepsilon_3$ 与 $X_2$ 不相关是无意义的,因为既然 $\varepsilon_3$ 对 $X_3$ 有直接因果作用,$\varepsilon_3$ 也必然直接作用于 $X_2$。这样一来,$\varepsilon_3$ 就通过 $X_3$ 成为了 $X_2$ 的间接原因,$cov(\varepsilon_3, X_2)$也就不可能等于 0。

# 第 2 章 | 可辨识性问题

在前面我们提到，所有递归模型都是可辨识的，因此也满足了获得有意义的参数估计值的一个关键的必要条件。如果所有递归模型都是可辨识的，那么是什么原因导致某些非递归模型不可辨识呢？凭直觉，似乎只要有优质且大样本的数据，我们就可以确定模型的参数。但看看最简化的非递归模型我们就会知道，这种直觉是错误的。请看下面的模型及其方程：




$$X_1 = \beta_{12} X_2 + \varepsilon_1 \quad [2.1]$$

$$X_2 = \beta_{21} X_1 + \varepsilon_2 \quad [2.2]$$

这个模型是不可辨识的。这里的不可辨识本质上就是大家都熟知的无法仅借助对一个时间点上 $X_1$ 和 $X_2$ 之间关系（例如相关）的了解来推导 $X_1$ 和 $X_2$ 之间的因果关系的方向。若允许变量之间互为因果关系，我们就不能简单地用一个参数来表示 $X_1$ 和 $X_2$ 之间关系的强度。我们必须找出这个关系中的哪一部分是沿着其中一个因果方向发挥作用，而另一部分是沿着另一个因果方向发挥作用的。在更复杂的非递归模型中，也会发生类似的问题。具体来说，在变量相同的情况下，非递归模型经常包含比递归模型更多的需要估计的参数。另外，与处理递归模型相比，研究者在处理非递归模型时，往往面对更少的信息，因为他们没有假定所有的误差项之间互不相关。

## 一个供给与需求的例子

为了理解可辨识性的意义，我们先来看一个涉及经济学中供给与需求概念

的解说性模型。介绍辨识性概念时，最常用的例子是关于确定供给曲线和需求曲线参数所需条件的讨论。

首先考虑某些农产品的需求曲线，比如小麦。需求曲线显示了消费者在任何给定价位所愿意购买的小麦总量。我们假定消费者的行为在整个研究期间充分相似，以满足需求曲线不随时间发生变化的假设。为了简化分析，我们还假定需求曲线是线性的，可以用下列方程来表示：

$$D_t = a_D + b_D P_t \qquad [2.3]$$

其中，$D_t$ 表示消费者在时间 t 愿意购买的小麦数量，$P_t$ 表示在同一时期单位小麦的价格，$a_D$ 和 $b_D$（分别是截距和斜率）则被假定为不随时间变化的常数。我们假设消费者愿意购买的小麦数量与价格成反比（这样假设具有典型性），从而缩小斜率的取值区间，使 $b_D<0$。基于这个假设我们可知，这一典型的需求曲线会如图 2.1 所示的那样下滑。如果能获得 $D_t$（消费者愿意购买的小麦数量）在两个不同价格（图 2.1 中的 $P_1$ 和 $P_2$）上的值，我们就能知道需求曲线上两个点（点 1 和点 2）的位置。既然两点确定一条直线，我们就能知道这条直线在图 2.1 中的位置，从而知道需求曲线的参数 $a_D$ 和 $b_D$ 的值。

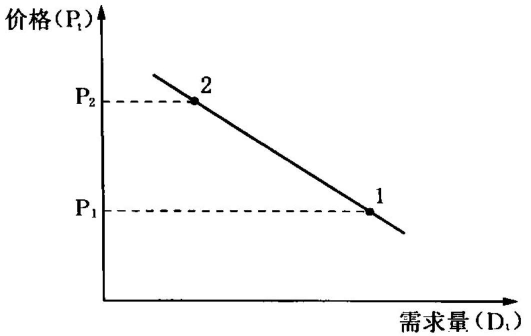



**图 2.1　方程 2.3 所示的需求曲线**

当然，我们还必须把供给曲线引入小麦市场模型中。供给曲线显示了小麦生产者在任意给定价位所愿意出售的小麦总量。与需求曲线一样，我们也假定供给曲线是线性的，且在研究期间具有稳定性。因此，它可以用下列方程来表述：

$$S_t = a_S + b_S P_t \qquad [2.4]$$

$S_t$ 代表生产者在时间 t 愿意出售的小麦数量，$a_S$ 和 $b_S$ 则假定为不随时间变化的常数。我们可以合理地假定价格和生产者愿意出售的小麦数量成正比，因此我们可以限定斜率值 $b_S$ 大于 0。与需求曲线的情况类似，如果能获得生产者在两个或多个价格上愿意出售的小麦数量，我们就能确定供给曲线上至少两个点的位置，从而确定供给曲线的参数 $a_S$ 和 $b_S$ 的值。

然而，在现实世界中，我们既无法获得消费者在多个价格上愿意购买的小麦数量，也无法获得生产者在多个价格上愿意出售的小麦数量。我们只能得到在某一个价格上的购买量和出售量，也就是说，农户愿意出售的小麦数量的价格等于消费者愿意购买的小麦数量的价格，即 $S_t = D_t$ 时的价格。既然在这个价格上 $S_t$ 确实等于 $D_t$，我们就可以规定 $S_t = D_t = Q_t$，其中 $Q_t$ 代表在时间 t 上小麦的出售量。因此，我们就可以把需求和供给曲线的方程 2.3 和方程 2.4 改成如下形式：

$$Q_t = a_D + b_D P_t \qquad [2.5]$$

$$Q_t = a_S + b_S P_t \qquad [2.6]$$

因此，在现实世界中，我们所能获得的是以实际价格 $p_o$ 出售的小麦的数量 $q_o$。这使我们能在图 2.2 中确定点 1 的位置，但没有足够的信息来确定供给或需求曲线上任何其他点的位置。我们仅仅知道需求曲线和供给曲线都经过点 1，除了 $b_D < 0$ 和 $b_S > 0$ 之外，我们对需求或供给曲线的参数特性一无所知。比如，

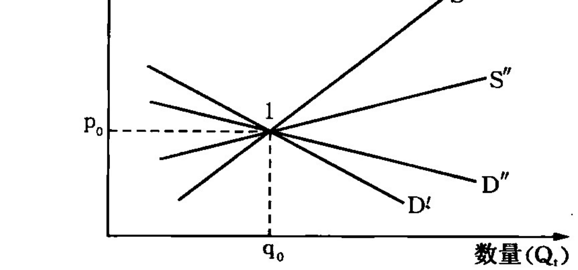



**图 2.2　方程 2.5 和方程 2.6 所示的需求和供给曲线**

我们的观察符合下述两种假设：(1)供给曲线是 $S'$，需求曲线是 $D'$；(2)供给曲线是 $S''$，需求曲线是 $D''$。但这样导致的结果是，我们无法揭示需求和供给曲线的位置，也就无法确定方程 2.5 和方程 2.6 的参数。在这种情况下，我们只能说方程 2.5 和方程 2.6 是不可辨识的，因为我们的观察与多组模型的参数值一致。[12]

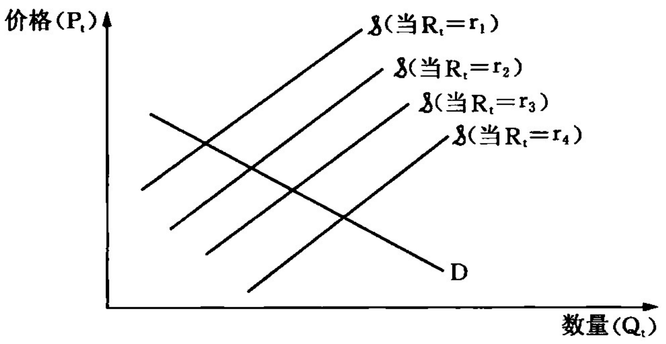
假设我们认为小麦的供给曲线会随时间而波动，而需求曲线（如以前那样）保持稳定。举例来说，我们可以认为，某一时期的降雨量是影响供给曲线而非需求曲线的外生变量。对于任一给定的价格，随着降雨量的增加，我们预期农作物也将增产，从而导致单位小麦生产成本的降低。随着小麦种植成本的下降，农民会相应增加在某一给定价格上的小麦供给量。相反，我们可以合理地假定某段时间内的降雨量并不会直接影响小麦消费者在给定价格上的购买量。

降雨量和生产者所愿意出售的小麦数量之间的关系可以用下述修正后的供给方程表示：

$$Q_t = a_S + b_S P_t + b_R R_t \qquad [2.7]$$

$R_t$ 表示时间段 t 内的降雨量，参数 $b_R$ 被假定大于 0。由于供给方程现在包含了 3 个变量，所以它就无法用一个二维的图来表示，它是在三维空间中的一个供给平面。图 2.3 通过呈现几个不同降雨量值（$r_1$、$r_2$、$r_3$ 和 $r_4$）的水平曲线，描述了这一供给平面（标记为 S）。供给平面上的每一条水平曲线都可被视为一条与特定降雨量相关的供给曲线，描述价格与生产者所愿意出售的小麦数量之间的关系。图 2.3 表明，随着降雨量从 $r_1$、$r_2$、$r_3$ 向 $r_4$ 逐渐增长，供给平面上的水平曲线不断向右下方移动（这里要注意的是，图 2.3 中不断变动的供给水平曲线的斜率是人为选择的，因为当水平曲线的斜率不受降雨量影响而保持不变时，它的斜率是未知的）。但是，当供给水平曲线因降雨量而发生变化时，需求曲线 D 仍处于固定位置。结果，固定的需求曲线和变动的供给水平曲线的交点形成了一条直线。因为这条直线上所有的点都在需求曲线上，所以由这些交点形成的这条直线就是需求曲线。因此，如果我们能在至少两个不同的降雨量期间观察到小麦的价格以及在这个价格上的小麦出售量，那么需求曲线 D 上至少有两个点可以被确定。这样，我们就能确定需求曲线的斜率和截距，即方程 2.5 中的参数

$a_D$ 和 $b_D$。所以，通过把降雨量这个外生变量引入模型，需求曲线（方程 2.5）成为可辨识的，因为借助小麦的价格和不同时期的小麦出售量可以求出方程中的参数。




注：$r_1 < r_2 < r_3 < r_4$。

**图 2.3　方程 2.3 所示的需求曲线 D 和方程 2.7 所示的供给平面 $\mathcal{S}$ 的层次曲线**

但在不同的时间点上（不同的降雨量水平），小麦的价格和出售量并不能帮助我们确定供给平面的有关参数。比如，平面上水平曲线的斜率依然是未知的，因为我们观察到的价格和售出量随时间变化的模式与方程 2.7 中的 $b_S$ 所取的任何正值都相符。因此，我们说方程 2.7 是不可辨识的。

如果我们对市场模型进行如下扩展，即假定人均可支配收入是一个影响消费者愿意购买小麦数量的外生变量，但不影响生产者愿意供给的小麦数量，我们就可以把需求方程修改成如下形式：

$$D_t = a_D + b_D P_t + b_I I_t \qquad [2.8]$$

$I_t$ 表示在时间 t 上的人均可支配收入，参数 $b_I$ 被假定为正值。对这一方程的分析类似于上述分析，在模型的需求方程中引入 $I_t$ 能够确定供给平面。随时间变动的需求平面上的水平曲线取决于可支配收入的水平，从而勾画出供给平面上的点。因此，如果我们能观察到两个或更多收入水平时的小麦价格和售出量，我们就有足够的信息来确定供给平面的相关参数。

这个例子说明，通过作出更强的假设（这些假设是关于消费者愿意购买的小麦数量 $D_t$ 与生产者愿意出售的小麦数量 $S_t$ 的影响因素的），有助于我们求出

小麦市场模型的方程。在其原初形式(方程 2.3 和方程 2.4)中,对于小麦价格和出售量之间关系的观测数量,并不能向我们提供充分的信息来确定模型方程中的参数。虽然增加观测数量是无用的,但增加有关 $D_t$ 和 $S_t$ 影响因素的假设却使我们成功地求出了方程。我们看到,在模型中引入外生变量 $I_t$(收入水平)和 $R_t$(降雨量)并作出下述两个限定,是供给和需求方程具有可辨识性的关键因素:(1)收入直接影响消费者愿意购买小麦数量的同时,并不直接影响生产者愿意出售的小麦数量;(2)降雨量直接影响生产者愿意出售的小麦数量的同时,并不直接影响消费者愿意购买的小麦数量。确实,假定某个特定的外生变量只影响模型中的某些而非全部的内生变量,是在社会科学研究中为了获得模型的可辨识性而经常使用的一种方法。

## 可辨识性与不可辨识性

一个理想的多方程模型应该是在观察足够充分的情况下,只有一组参数能够同时符合观测数据和模型的限定要求。但在最初的小麦市场模型(方程 2.3 和方程 2.4)中,我们遇到的问题是,小麦价格和出售量这两个变量的观测信息并不足以用来确定模型的参数。简言之,这正是模型的不可辨识问题。在这种情况下,即使拥有最完备的经验数据,我们的观测数据依然与一组以上的参数相符。当然,模型的一组参数可被看做是对模型所描述过程的一种解释,我们的观测可能与多种解释相符。因此,当面对一个不可辨识的模型时,对于变量之间经验关系的进一步观测并不能解决这一问题。为了辨识一个不可辨识的模型,我们必须对模型中的变量进行更进一步的先验假设,通过对其形成足够的限制,使只有一个解释(即一组参数)能同时符合先验的假设限定和变量之间的经验关系。

迄今为止,当讨论"可辨识性"的意义时,我们采用了这样的陈述:"如果充分或完备的观测数据能决定方程的唯一参数,那么该方程就是可辨识的。"为了更加精确地定义"可辨识性"这一概念,我们需要澄清"充足或完备"的观测的含义。首先,我们要明白缺乏辨识性不是一个统计推论问题。一个模型永远不会

因为缺乏足够的样本而不能被辨识，也不会因为与抽样有关的问题而不可辨识。因此，在定义可辨识性时，假定我们有完备的观测是有用的，即拥有在模型所有的外生变量取值上的内生变量条件分布的完备信息。如果这一条件概率分布能唯一地决定方程的参数，那么我们才能说该多方程模型是可辨识的。

Christ(1966:第 8 章)把能够产生针对外生变量的内生变量完备条件分布的数据称做一个“恰当选择的无限大样本”。这个假设中的无限样本是指一个包含外生变量每一种可能的取值组合的样本，并且该样本能够为每个取值组合提供无限多个观测数据。利用这一概念，我们可以从另一个角度来定义可辨识性。对于一个恰当选择的无限大样本而言，如果所获得的数据足以唯一地决定某方程的一组参数，那么该方程就是可辨识的。而如果该方程是可辨识的，就只能有一个解释(即一组参数)同时符合来自无限大样本的数据和模型的假设限定。显然，无限大样本数据是不可能在有限次数的观测中获得的，因此无限大样本这一概念完全是假设出来的。但利用这一概念，我们却能澄清某方程是否可辨识，这与由从一个有限样本到另一个有限样本的数据变异而引起的任何统计推论问题完全无关。相反，如果一个针对外生变量的内生变量的条件概率分布不能唯一地决定一组参数值，也就是说，存在多组参数值契合概率分布，我们就说该方程是不可辨识的。换言之，如果一个选择恰当的无限大样本的数据存在一组以上与之相契合的参数值，那么该方程就是不可辨识的。显而易见，如果一个无限大样本的数据不足以提供充分的信息来决定方程的参数，那么尝试用现实中的有限样本来估计方程的参数必然是徒劳的。所以，只要模型中的某个方程是不可辨识的，就不可能获得有意义的方程参数估计值。

因此，面对一个包含不可辨识方程的非递归模型，在进行参数估计之前，有必要先修改模型以使方程变成可辨识的。要使一个不可辨识的方程变得可辨识，必须设定先验假设，以进一步限定模型中的方程。这种限定形式是多样的。其中一种形式是假设模型中的一对参数相等或具有已知的比例关系。另外，对模型误差项分布的多种限定方法，有时候也足以把不可辨识方程改造成可辨识的方程。[13]然而在实际运用中，辨识非递归模型中的方程最常用的先验假设是所谓的零限定。它假定模型中某些参数为 0，也就是说，模型中的某些变量对另

一些变量没有直接的因果作用。在各种各样的“辨识限定”中，本书（除附录 2 外）将严格局限于零限定的讨论，因为这是社会科学研究中唯一通用的方法。

在此小结一下我们所用的术语，我们说一个多方程模型是可辨识的，仅当其中的每一个方程都是可辨识的；说一个多方程模型是不可辨识的，当其中的任何一个方程都是不可辨识的。总而言之，不可辨识不是测量或数据质量的问题。即使拥有最好的数据，比如对于一个大样本（甚至是无限大）而言有效且可信的测量指标，不可辨识性仍然无法被克服。因此，不可辨识性不是一个统计问题，而是一个模型设定问题，在逻辑上先于通过样本数据所进行的参数估计。可辨识性其实是一个假设性问题，如果选择恰当的无限大样本的数据可获得（当然这在现实中是不可能的），那么一组唯一的参数能否确定？确实，不可辨识并不意味着参数估计不可行。某人可以为一个不可辨识的方程收集一组样本数据，然后使用一种统计技术——比如一般最小二乘法回归分析——获得参数估计，但问题是，不可辨识性使任何参数估计都毫无意义。

## 适度可辨识性和过度可辨识性

如果一个方程是不可辨识的，我们就必须给模型添加更进一步的限定，使其在获得恰当选择的无限大样本数据的情况下能唯一地确定一组参数。这样，方程就变得可辨识了。在这些可辨识的方程中，我们还要区分适度辨识的方程和过度辨识的方程。如果添加在模型上的限定恰好可使方程被辨识，该方程就是适度辨识的，即在这些限定条件下，方程可被辨识，但在这些限定的任何子集中，方程都不可辨识。相反，如果模型的限定超过恰好使方程可被辨识的一组限定，我们就说该方程过度辨识。[14]

其实，如果真能获得一组来自一个选择恰当的无限大样本的数据，就没必要区分过度辨识和适度辨识的方程。因为对于这两类方程中的任何一类，无限大样本的数据都能确定一组唯一的参数。但在现实中，我们只能利用有限的样本来估计参数，因此过度辨识和适度辨识方程的区分依然很重要。对于来自一个有限样本数据的适度辨识方程，我们可以运用一定的方法来确定一组唯一的

参数估计。但对于一个过度辨识方程，不可能获得一组参数估计既与一个有限样本数据相契合，又满足模型所有的先验假设限定。如果我们放弃一个或多个模型限定，一个过度辨识方程就能转变为适度辨识方程，从而产生一组唯一的参数估计 $S_1$。但如果我们放弃了另一组模型限定，使方程适度被辨识，那么就会产生一组不同的唯一参数估计 $S_2$。可见，对于一个过度辨识方程，放宽不同的模型假设会产生不同的参数估计。对这样的方程进行参数估计时，就会产生这样一个统计问题：需要弄清楚因为放宽不同的模型限定而产生的多组参数估计之间的差异，从而获得一组合理的参数估计。

## 为什么某些非递归模型无法被辨识？

我们看到，当一个方程不可辨识时，从一个恰当选择的无限大样本获得的信息不足以确定一组唯一的参数估计。为了更深入细致地了解为什么这样的数据对于某些非递归模型来说不够充分，我们将以几个解说性模型为例，从简化式视角和线性组合视角两方面来进行分析。这两种分析视角有助于我们了解使非递归模型中的方程变得可辨识的必要条件，也有助于提出一种用于检测模型可辨识性的策略，这一策略将被整合到第 3 章所要讨论的检测程序之中。

### 简化式视角

为了解说方便起见，我们首先以投票行为的分层非递归模型为例，该模型基于图 1.2 及方程 1.3 和方程 1.4 所表述的因果关系，其方程形式如下：

$$X_1 = \gamma_{13} Z_3 + \varepsilon_1 \qquad [2.9]$$

$$X_2 = \beta_{21} X_1 + \gamma_{23} Z_3 + \varepsilon_2 \qquad [2.10]$$

因为上述方程与方程 1.3 和方程 1.4 一样，所以该模型本质上依然是分层模型。但我们将放宽误差项互不相关这一不现实的假设，而代之以另一个在第 1 章提及的非递归模型中更为典型的误差项假设：(1)误差项与所有外生变量不相关，即 $[\text{cov}(\varepsilon_1, Z_3) = \text{cov}(\varepsilon_2, Z_3) = 0]$；(2)为便利起见，假定的 $E(\varepsilon_1) =$

$E(\varepsilon_2)=E(X_1)=E(Z_3)=0$。

要确定某个非递归模型中的方程是否可辨识的一种方法，是检验与模型的结构方程相对应的简化式方程。在结构方程形式中，每个内生变量被表述为对其有直接因果作用的变量(外生变量和内生变量)加上一个误差项的函数，如方程2.9和方程2.10。如果我们改写模型的结构方程，将每个内生变量表述为仅仅是外生变量和一个误差项的函数，这样，每个方程仅在等号左边存在内生变量，这就是方程的简化式。要把结构方程转变为简化式，需要把等号右边的内生变量替换成其在结构方程中的表达式。比如，要把定义 $X_2$ 的方程2.10转换成简化式，就需要用 $\gamma_{13}Z_3+\varepsilon_1$ 来替换等号右边的 $X_1$，得到方程：

$$X_2=\beta_{21}(\gamma_{13}Z_3+\varepsilon_1)+\gamma_{23}Z_3+\varepsilon_2 \qquad [2.11]$$

把方程式右边的各项乘出来：

$$X_2=\beta_{21}\gamma_{13}Z_3+\beta_{21}\varepsilon_1+\gamma_{23}Z_3+\varepsilon_2 \qquad [2.12]$$

最后，再合并同类项就得到了简化式方程：

$$X_2=(\gamma_{23}+\beta_{21}\gamma_{13})Z_3+(\beta_{21}\varepsilon_1+\varepsilon_2) \qquad [2.13]$$

再看方程2.9，由于等号右边没有内生变量，所以已经是简化式了。因此，方程2.9和方程2.13就构成了模型的简化式方程组。最后，如果我们为方程组中的参数重新进行命名，如下：

$$\pi_{13}=\gamma_{13} \qquad [2.14]$$

$$U=\varepsilon_1 \qquad [2.15]$$

$$\pi_{23}=\gamma_{23}+\beta_{21}\gamma_{13} \qquad [2.16]$$

$$V=\beta_{21}\varepsilon_1+\varepsilon_2 \qquad [2.17]$$

那么，简化式方程就可被表述为：

$$X_1=\pi_{13}Z_3+U \qquad [2.18]$$

$$X_2=\pi_{23}Z_3+V \qquad [2.19]$$

其中，U 和 V 都与 $Z_3$ 不相关。[15]

虽然简化式和结构方程都是同一个模型的表达式，但它们之间存在区别。简化式方程的参数仅仅告诉我们外生变量每变化一个单位所引起的内生变量的变化量。相反，结构方程反映的是模型所表示的过程的内在因果联系，从而告诉我们为什么外生变量每变化一个单位，内生变量就会产生一定量的变化。另外，在可辨识性上，两类形式的方程也存在差异。一个非递归模型的简化式方程必然是可辨识的，因为方程式右边的解释变量与模型中的误差项都不相关。

非递归模型简化式方程的这种可辨识的必然性有什么意义？我们来看方程 2.9 和方程 2.10 所代表的模型(其简化式是方程 2.18 和方程 2.19)。由于两个简化式的方程都是可辨识的，如果我们能获得一个来自选择恰当的无限大样本的数据，$\pi_{13}$ 和 $\pi_{23}$ 的唯一取值就能被确定。但我们的最终目标是确定结构方程中的参数值 $\gamma_{13}$、$\gamma_{23}$ 和 $\beta_{21}$。为此，一个可行的策略是用方程 2.14 到方程 2.17 来替代简化式参数 $\pi_{13}$ 和 $\pi_{23}$，然后求解出结构方程参数。倘若我们成功了，也因此获得了一个特定结构方程每个参数的唯一解，那么就说明这个方程是可辨识的。例如，有关简化式参数 $\pi_{13}$ 和 $\pi_{23}$ 的信息就足以求出结构方程 2.9 中的参数 $\gamma_{13}$，因为通过求解方程 2.14 到方程 2.17，我们得到 $\gamma_{13}$ 等于 $\pi_{13}$。因此，方程 2.9 是可辨识的。但另一方面，通过求解方程 2.14 到方程 2.17，我们并不能得到 $\gamma_{23}$ 和 $\beta_{21}$ 的唯一解，因为这两个参数存在无穷个解。这表明方程 2.10 是不可辨识的。

一般来说，多方程模型中的某个结构方程是否可辨识的问题可简化为，如果确切地知道模型简化式的参数，是否存在足够的信息来确定结构方程的参数？如果结构方程的参数能且仅能通过模型的简化式参数获得唯一解，该结构方程就是适度可辨识的；如果简化式参数不足以获得结构方程参数的唯一解，该结构方程就是不可辨识的。当然，对拥有两个以上内生变量的非递归模型，为确定简化式参数能否求出结构方程参数的唯一解，进行类似方程 2.14 到方程 2.17 的等式运算无疑是非常繁复沉闷的，这就体现了第 3 章中辨识检测的价值。

简化式视角也可用来确定某个不分层的非递归模型中的方程是否可辨识。为了便于说明，我们引入 Kritzer(1977)关于抗议事件中暴力行为的模型作为例

子，并进行了适当的修改（在此，我们剔除了 Kritzer 模型的几个变量，而在第 5 章中，我们将讨论 Kritzer 的完全模型）。图 2.4 的因果图示显示了这个简化的抗议事件中的暴力行为模型，其方程如下：

$$X_1 = \beta_{12}X_2 + \gamma_{13}Z_3 + \varepsilon_1 \qquad [2.20]$$

$$X_2 = \beta_{21}X_1 + \gamma_{23}Z_3 + \gamma_{24}Z_4 + \varepsilon_2 \qquad [2.21]$$

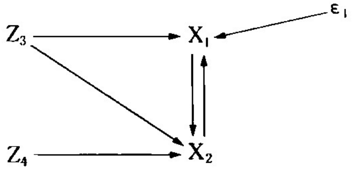



**图 2.4　方程 2.20 和方程 2.21 所示的非递归模型的因果图示**

其中，$X_1$ 表示抗议者的暴力水平，$X_2$ 表示警察的暴力水平。这两个内生变量被假定受到外生变量 $Z_3$（抗议者所遵奉的公民反抗特性）的直接影响。最后，警察拥有重装备的程度被看做是一个仅直接影响警察暴力的外生变量（$Z_4$），它对抗议者暴力只有间接影响。

把方程 2.20 化为简化式，我们把定义 $X_2$ 的方程 2.21 代入方程 2.20：

$$X_1 = \beta_{12}(\beta_{21}X_1 + \gamma_{23}Z_3 + \gamma_{24}Z_4 + \varepsilon_2) + \gamma_{13}Z_3 + \varepsilon_1 \qquad [2.22]$$

如果我们对 $X_2$ 的结构方程也进行类似的变换，然后对获得的方程以及方程 2.22 进行适当的同类项合并，就可以得到下列方程以表述图 2.4 抗议事件中暴力行为模型的简化式方程：

$$X_1 = \frac{(\gamma_{13} + \beta_{12}\gamma_{23})Z_3 + \beta_{12}\gamma_{24}Z_4 + (\varepsilon_1 + \beta_{12}\varepsilon_2)}{1 - \beta_{12}\beta_{21}} \qquad [2.23]$$

$$X_2 = \frac{(\gamma_{23} + \beta_{21}\gamma_{13})Z_3 + \gamma_{24}Z_4 + (\varepsilon_2 + \beta_{21}\varepsilon_1)}{1 - \beta_{12}\beta_{21}} \qquad [2.24]$$

然后，这两个简化式可被写为：

$$X_1 = \pi_{13}Z_3 + \pi_{14}Z_4 + U \qquad [2.25]$$

$$X_2 = \pi_{23} Z_3 + \pi_{24} Z_4 + V \qquad [2.26]$$

其中,

$$\pi_{13} = \frac{\gamma_{13} + \beta_{12}\gamma_{23}}{1 - \beta_{12}\beta_{21}} \qquad [2.27]$$

$$\pi_{14} = \frac{\beta_{12}\gamma_{24}}{1 - \beta_{12}\beta_{21}} \qquad [2.28]$$

$$U = \frac{\varepsilon_1 + \beta_{12}\varepsilon_2}{1 - \beta_{12}\beta_{21}} \qquad [2.29]$$

$$\pi_{23} = \frac{\gamma_{23} + \beta_{21}\gamma_{13}}{1 - \beta_{12}\beta_{21}} \qquad [2.30]$$

$$\pi_{24} = \frac{\gamma_{24}}{1 - \beta_{12}\beta_{21}} \qquad [2.31]$$

$$V = \frac{\varepsilon_2 + \beta_{21}\varepsilon_1}{1 - \beta_{12}\beta_{21}} \qquad [2.32]$$

通过方程 2.27 到方程 2.32 能够求出 $\beta_{12} = \pi_{14}/\pi_{24}$ 以及 $\gamma_{13} = \pi_{13} - (\pi_{14}\pi_{23}/\pi_{24})$,因此该模型的简化式参数 $\pi_{13}$、$\pi_{14}$、$\pi_{23}$ 和 $\pi_{24}$ 能够确定方程 2.20 的结构参数 $\beta_{12}$ 和 $\gamma_{13}$ 的唯一解。这样一来,图 2.4 模型中 $X_1$ 的结构方程就是可辨识的,但通过方程 2.27 到方程 2.32,我们无法求出方程 2.21 的结构参数 $\beta_{21}$、$\gamma_{23}$ 和 $\gamma_{24}$ 的唯一解,因此 $X_2$ 的方程是不可辨识的。

以简化式视角来看上述两个模型,其中所有可辨识的方程都是适度辨识的。因此,最后有必要举一个过度辨识的模型的例子。我们所要引用的这个模型是基于 Duncan、Haller 和 Portes(1971)对青少年的同伴互动如何影响其职业和教育期望的研究。我们将对他们的原初模型进行简化,剔除与教育期望水平相关的变量。随后,我们会分析一个与 Duncan 等人的原初模型非常接近的模型。我们的职业期望模型如图 2.5 所示,其方程形式如下:

$$X_1 = \beta_{12} X_2 + \gamma_{13} Z_3 + \gamma_{14} Z_4 + \varepsilon_1 \qquad [2.33]$$

$$X_2 = \beta_{21} X_1 + \gamma_{25} Z_5 + \varepsilon_2 \qquad [2.34]$$

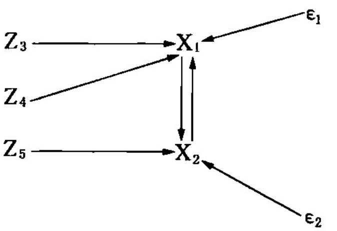



**图 2.5 方程 2.33 和方程 2.34 所示的非递归模型的因果图示**

模型中内生变量分别是 $X_1$（男性青少年的职业期望水平）和 $X_2$（其友人的职业期望水平），假定它们互为因果关系。对于受访者及其朋友而言，智力（$Z_3$ 表示受访者的智力，$Z_5$ 表示他朋友的智力）被假定为一个影响职业期望的外生变量。同时，我们还假定能获得受访者的父母对受访者职业期望的看法（比如，受访者在多大程度上认为他父母鼓励他争取高水平成就），这是一个对受访者职业期望水平有直接影响的外生变量（$Z_4$）。通过替换参数和合并同类项，得到了下面的模型简化式方程：

$$X_1 = \frac{\gamma_{13}Z_3 + \gamma_{14}Z_4 + \beta_{12}\gamma_{25}Z_5 + (\varepsilon_1 + \beta_{12}\varepsilon_2)}{1 - \beta_{12}\beta_{21}} \qquad [2.35]$$

$$X_2 = \frac{\beta_{21}\gamma_{13}Z_3 + \beta_{21}\gamma_{14}Z_4 + \gamma_{25}Z_5 + (\varepsilon_2 + \beta_{21}\varepsilon_1)}{1 - \beta_{12}\beta_{21}} \qquad [2.36]$$

这两个方程可以被写为下列形式：

$$X_1 = \pi_{13}Z_3 + \pi_{14}Z_4 + \pi_{15}Z_5 + U \qquad [2.37]$$

$$X_2 = \pi_{23}Z_3 + \pi_{24}Z_4 + \pi_{25}Z_5 + V \qquad [2.38]$$

其中，

$$\pi_{13} = \frac{\gamma_{13}}{1 - \beta_{12}\beta_{21}} \qquad [2.39]$$

$$\pi_{14} = \frac{\gamma_{14}}{1 - \beta_{12}\beta_{21}} \qquad [2.40]$$

$$\pi_{15} = \frac{\beta_{12}\gamma_{25}}{1 - \beta_{12}\beta_{21}} \qquad [2.41]$$

$$U = \frac{\varepsilon_1 + \beta_{12}\varepsilon_2}{1 - \beta_{12}\beta_{21}} \qquad [2.42]$$

$$\pi_{23} = \frac{\beta_{21}\gamma_{13}}{1 - \beta_{12}\beta_{21}} \qquad [2.43]$$

$$\pi_{24} = \frac{\beta_{21}\gamma_{14}}{1 - \beta_{12}\beta_{21}} \qquad [2.44]$$

$$\pi_{25} = \frac{\gamma_{25}}{1 - \beta_{12}\beta_{21}} \qquad [2.45]$$

$$V = \frac{\varepsilon_2 + \beta_{21}\varepsilon_1}{1 - \beta_{12}\beta_{21}} \qquad [2.46]$$

现在假设已知简化式的参数，求解方程组的结构参数，我们会发现从方程 2.40 和方程 2.44 中可以求出 $\beta_{21} = \pi_{24}/\pi_{14}$，而从方程 2.39 和方程 2.43 中则得出 $\beta_{21} = \pi_{23}/\pi_{13}$。如果方程 2.33 和方程 2.34 是对所要研究过程的精确模型化，而且能获得“恰当选择的无限大样本”数据，那么存在两个 $\beta_{21}$ 的表达式并没有问

题，因为两个表达式会得到一个相同的 $\beta_{21}$ 值。这样我们就得到：

$$\beta_{21}=\frac{\pi_{24}}{\pi_{14}}=\frac{\pi_{23}}{\pi_{13}} \qquad [2.47]$$

但实际情况是，真正的简化式参数是不可知的，因为不存在无限大的样本。我们只能通过有限样本数据来估计简化式参数。最常用的方法是对方程 2.37 和方程 2.38 运用一般最小二乘法回归。但利用有限样本数据，我们就不能保证简化式的参数估计满足方程 2.47。我们的最大希望是它们近似相等，即：

$$\frac{\hat{\pi}_{24}}{\hat{\pi}_{14}}\simeq\frac{\hat{\pi}_{23}}{\hat{\pi}_{13}}$$

符号$\simeq$表示“近似等于”，而参数上的“帽子”(^)用来标记参数的估计值。因此，在实际运算中，当用简化式参数估计来替代方程中真正的 $\pi_{ij}$ 时，我们就无法获得任何一组满足方程 2.39 到方程 2.46 的结构参数估计值。方程 2.39 到方程 2.46 反而提供了两个有关 $\hat{\beta}_{21}$ 的表达式，即 $\hat{\beta}_{21}=\hat{\pi}_{24}/\hat{\pi}_{14}$ 和 $\hat{\beta}_{21}=\hat{\pi}_{23}/\hat{\pi}_{13}$。我们希望这两个方程会产生近似的估计值，但由于抽样的变异性，它们不可能相同。这就留下了一个有待解决的统计问题：如何从多个不一致的估计值中确定总体参数 $\beta_{21}$ 的一个合理估计值？

任何存在过度辨识方程的模型都会产生类似的问题。如果存在超过“最小限度充分”辨识的冗余限定，在利用简化式参数求解结构参数的过程中，结构参数会产生多重解。这样，当我们把简化式参数的估计值代入多重解之中时，就产生了多个不一致的结构参数估计值。而当模型中的所有方程都是适度辨识的，上述情况就不会发生。因为利用简化式参数来求解结构参数，只能获得唯一解，因而每个结构参数就只有一个表达式。当把简化式参数的估计值代入这些表达式中，我们就得到了每个结构参数的唯一估计值。

### 线性组合视角

另一个检验多方程模型是否可辨识的便利方法是线性组合视角。一组方程的线性组合就是把其中的每个方程乘以一个常数项(每个方程可以不同)，

然后相加得到的一个方程。比如，$38x+2y=82$ 是下述方程的一个线性组合：(1) $4x+2y=8$；(2) $3x+5y=17$；(3) $5x-6y=13$。如果将方程(1)乘上 7，方程(2)乘上 0，方程(3)乘上 2，我们就能得到：(1) $28x+14y=56$；(2) $0=0$；(3) $10x-12y=26$。把这 3 个等式相加就得到了目标方程：$38x+2y=82$。我用"非零"线性组合来指代一组方程的线性组合，其中任何乘法运算中的常数项都不为 0。

借助这些定义，我们假定有一个以结构方程的形式呈现的多方程模型满足方程组 1.2 的假设。这样一来，这个模型中方程的线性组合就具有一些非常重要的特性。首先，任何包含方程的模型，如果每个方程都是其原初模型中结构方程的一个线性组合，那么这个模型必然有与原初模型相同的简化式方程。反之亦然，那些与原初模型具有相同简化式方程的模型，就是那些其所包含的方程为原初模型方程的线性组合的模型。其次，任何一个通过对原初模型的结构方程组进行线性组合而得到的方程，都必然与原初模型的选择恰当的无限大样本数据相吻合。反之亦然，任何与该无限大样本数据相吻合的方程都必然是原初模型结构方程组的一个线性组合。[16]这些特性使我们能够确立一项任何可辨识方程都要满足的条件：对于多方程模型中的某个方程而言，只有当模型中结构方程的所有线性组合(除该方程自身外)都无法满足针对该方程的限定时①，该方程才是可辨识的。

为了清楚地说明这一条件，让我们再次检验图 2.4 所示的抗议事件暴力行为模型，其表达式为方程 2.20 和方程 2.21。首先将方程 2.20 乘以一个非零常数 λ，方程 2.21 乘以非零常数 μ，然后把它们相加。与常数相乘后，我们得到：

$$\lambda X_1 = \lambda\beta_{12}X_2 + \lambda\gamma_{13}Z_3 + \lambda\varepsilon_1 \quad [2.48]$$

$$\mu X_2 = \mu\beta_{21}X_1 + \mu\gamma_{23}Z_3 + \mu\gamma_{24}Z_4 + \mu\varepsilon_2 \quad [2.49]$$

把方程 2.48 和方程 2.49 相加，再合并同类项得到：

$$0 = (\lambda\beta_{12} - \mu)X_2 + (\mu\beta_{21} - \lambda)X_1 + (\lambda\gamma_{13} + \mu\lambda_{23})Z_3 + (\mu\gamma_{24})Z_4 + (\lambda\varepsilon_1 + \mu\varepsilon_2) \quad [2.50]$$

① 即提供与该方程完全相同的变量。——译者注

然后，方程两边都减去 $(\lambda\beta_{12}-\mu)X_2$，再除以 $-(\lambda\beta_{12}-\mu)$，就得到下面这个关于方程 2.20 和方程 2.21 的线性组合：

$$X_2=-\frac{\mu\beta_{21}-\lambda}{\lambda\beta_{12}-\mu}(X_1)-\frac{\lambda\gamma_{13}+\mu\gamma_{23}}{\lambda\beta_{12}-\mu}(Z_3)-\frac{\mu\gamma_{24}}{\lambda\beta_{12}-\mu}(Z_4)-\frac{\lambda\varepsilon_1+\mu\varepsilon_2}{\lambda\beta_{12}-\mu} \qquad [2.51]$$

这个方程可以换成下列形式：

$$X_2=\beta_{21}^*X_1+\gamma_{23}^*Z_3+\gamma_{24}^*Z_4+\varepsilon_2^* \qquad [2.52]$$

其中，

$$\beta_{21}^*=-\frac{\mu\beta_{21}-\lambda}{\lambda\beta_{12}-\mu}$$

$$\gamma_{23}^*=-\frac{\lambda\gamma_{13}+\mu\gamma_{23}}{\lambda\beta_{12}-\mu}$$

$$\gamma_{24}^*=-\frac{\mu\gamma_{24}}{\lambda\beta_{12}-\mu}$$

$$\varepsilon_2^*=-\frac{\lambda\varepsilon_1+\mu\varepsilon_2}{\lambda\beta_{12}-\mu}$$

最后，方程 2.52 中的误差项($\varepsilon_2^*$)的均值为 0，且与外生变量 $Z_1$ 和 $Z_2$ 不相关。因此，方程 2.51 在形式上就与 $X_2$ 的原初结构方程完全一样(包含相同的变量 $X_1$、$X_2$、$Z_3$ 和 $Z_4$)。

方程 2.21 和方程 2.51 的相同形式表明，抗议事件暴力行为模型的简化式参数并不能使方程 2.21 的结构参数获得唯一解。这是为什么？我们知道，由方程 2.20 和方程 2.21 组成的模型与由方程 2.20 和方程 2.52 所组成的模型具有相同的简化式方程，因为方程 2.52 是方程 2.20 和方程 2.21 的线性组合。所以，如果已知抗议事件暴力行为模型的简化式参数来求解模型的结构参数，我们就会发现两组参数($\beta_{21}$、$\gamma_{23}$ 和 $\gamma_{24}$ 与 $\beta_{21}^*$、$\gamma_{23}^*$ 和 $\gamma_{24}^*$)被证明与简化式参数相吻合。这样一来，方程 2.21 就是不可辨识的。这一分析逻辑也可以换一种说法来表述。假定对于由方程 2.20 和方程 2.21 组成的抗议事件暴力行为模型，我们能够获得一个恰当选择的无限大样本数据。该数据也不能使我们获得方程

2.21 的结构参数的唯一解，因为只要参数 $\beta_{21}$、$\gamma_{23}$ 和 $\gamma_{24}$ 与数据吻合，那么参数 $\beta_{21}^{*}$、$\gamma_{23}^{*}$ 和 $\gamma_{24}^{*}$ 也同样会吻合。实际上还存在无穷个由方程 2.20 和方程 2.21 组成的非零线性组合的参数组与这样的数据相吻合。所以，我们依然说方程 2.21 是不可辨识的。就本质而言，是有可能得到抗议事件暴力行为模型方程的一个线性组合，其满足方程 2.21 的所有限定(即包含与方程 2.21 完全一样的变量)，这足以证明方程 2.21 是不可辨识的。

那么，模型中的结构方程 2.20 又如何？与先前的情形相反，如果(正如所假定的那样)$\gamma_{24}$ 不能等于 0，任何方程 2.20 和方程 2.21 所形成的非零线性组合都必然给 $Z_4$ 形成一个非零参数(比如，$\mu\gamma_{24}$)。因此，当方程 2.20 和方程 2.21 的所有非零线性组合与模型的选择恰当的无限大样本数据吻合时，其中任何一个线性组合都无法满足方程 2.20 的限定，即 $Z_4$① 项的系数为 0。因此，一个恰当选择的无限大样本数据足以确定方程 2.20 的结构参数的唯一解，这就确保了方程 2.20 是可辨识的。

通常来说，检验多方程模型中的某个结构方程的是否可辨识，往往是分析能否通过模型中部分或所有结构方程构成的线性组合，来获得一个与所要求解的方程形式相同的(即包含完全相同的变量)其他方程。如果找不到这样的方程，那么该方程就是可辨识的。反之，如果能找到这样一个由模型中部分或所有方程形成的线性组合(除了被研究的方程本身)，它恰好包含与所要求解的方程完全相同的变量，我们就可以确信该被研究的方程是不可辨识的。这表明，若模型中越多变量被假定在一个给定的方程中具有非零的参数(或者说，越多的变量离开了方程)，该方程可辨识的可能性就越大。因为参数为 0 的限定很难形成在形式上与这一给定方程完全一致的模型方程的线性组合。在我们进入第 4 章(关于如何把不可辨识的非递归模型修改成可辨识模型的议题)之前，这是一个必须牢记的关键法则。

---

① 原文中这里为 $Z_2$，但根据上下文，似应改为 $Z_4$。——译者注

# 第 3 章 | 可辨识性的检验

在上文中我们看到，简化式或线性组合方法都可以用来检验某个非递归模型是否可辨识。这两种方法在数学上是等价的，能够获得关于模型中的方程是否可辨识的相同结论。它们还提供了关于可辨识性的必要且充分条件，即每种方法都产生了一个保证某个方程是不可辨识的条件和一个保证某个方程是可辨识的条件。但从实际运用的角度来看，这两种方法即使对最简单的非递归模型，都是相当繁复和费时的。可喜的是，还存在另外两种更易操作的方法来检验可辨识性，其中一种是秩条件，它对可辨识性来说也是必要且充分的。在本章中，我将介绍一种来自秩且在数学上与秩等价的运算法则。另一种检验可辨识性常用的方法是次序条件，其优点是运用起来非常简便，但缺点在于，它只是可辨识性的一个必要但不充分条件。

## 次序条件

假定一个非递归模型的表达式为一组结构方程。我们设定：

m = 模型中内生变量的数目

k = 模型中外生变量的数目

我们依然假定该模型的所有变量和误差项的均值为 0[17]，而且另一个关于误差项的假定是每个误差项与所有的外生变量都不相关。在此，我想再一次提醒读者，对于一个特定的模型，如果关于误差项的假定不合理，或者读者想要作出一个更强的关于误差项的假定，那么本章所介绍的可辨识性的检验方法就不适

用，除非对其进行修改(参见附录 2)。

检验模型可辨识性最常用的一种方法是简单计数法，即次序条件。借助这个方法，模型中每个方程的检验都可以单独进行。假定我们要检验一个结构方程，为了使该方程可辨识，方程所未包含的外生变量数必须大于或等于方程所包含的作为解释变量的内生变量数。因此，如果我们设定：

$k_e$ = 待检验的结构方程所未包含的模型外生变量数

$m_i$ = 待检验的方程所包含的模型内生变量数(包括等式左边被解释的那个变量)

如果次序条件显示该方程是可辨识的，就必然满足下列条件：

$$k_e \geqslant m_i - 1 \quad [3.1]$$

右边减去 1 是因为 $m_i$ 既包含了方程中的解释性内生变量，也包含了被解释的内生变量。

用一些简单的加减法就能得到一个等价的次序条件计算式，我们设定：

$m_e$ = 待检验的结构方程所未包含的模型内生变量数

然后，我们在方程 3.1 的两边都加上 $m_e$，就得到：

$$m_e + k_e \geqslant m_e + m_i - 1 \quad [3.2]$$

而 $m_e + m_i = m$ (模型中内生变量的总数)，因此方程 3.2 就变为：

$$m_e + k_e \geqslant m - 1 \quad [3.3]$$

因此，对一个可辨识的方程而言，它所未包含的模型变量数(包含内生和外生变量)必须等于或大于模型的结构方程数目①减去 1。

这两种次序条件方法有利于我们迅速判断某些方程是否可辨识。如果某个方程不能满足次序条件，那么它必然是不可辨识的。但我们也必须清醒地认识到次序条件法的局限：它对可辨识性而言是必要条件，但不是充分条件。因此，某方程满足了次序条件并不足以证明它是可辨识的。在假定该方程可辨识

① 模型的结构方程数目即模型所包含的内生变量的数目。——译者注

(即能获得合理的参数估计)之前，还必须对该方程进行秩条件检验。

## 秩条件

秩条件是可辨识性的必要且充分条件，它采纳运用线性组合的视角来检验某个方程的可辨识性的程序。虽然介绍秩条件需要用到线性代数的知识，但本章将介绍一种在数学上与秩条件等价的运算法则，即使读者不具备线性代数知识，也能使用这个法则。[18]

该运算法则的基本原理就是用一个 $\xi$ 行的矩阵来表示模型中变量的系数，其中每一行对应模型的一个结构方程。以图 2.4 所示模型为例，如果我们改写结构方程 2.20 和方程 2.21 如下：

$$0 = -X_1 + \beta_{12}X_2 + \gamma_{13}Z_3 + \varepsilon_1 \quad [3.4]$$

$$0 = \beta_{21}X_1 - X_2 + \gamma_{23}Z_3 + \gamma_{24}Z_4 + \varepsilon_2 \quad [3.5]$$

就可以用下述矩阵来表示这两个结构方程：

$$\begin{array}{c} \\ X_1 \\ X_2 \end{array} \begin{array}{c} \begin{array}{cccc} X_1 & X_2 & Z_3 & Z_4 \end{array} \\ \begin{bmatrix} -1 & \beta_{12} & \gamma_{13} & 0 \\ \beta_{21} & -1 & \gamma_{23} & \gamma_{24} \end{bmatrix} \end{array}$$

由此可知，根据变量是否被方程所包含，我们可以把每个方程的系数在矩阵中简化为 0 或星号( * )，从而保留方程可辨识性检验的所有信息。这样，通过 0 和星号的操作。我们可以模拟下列运算：如果把一个方程乘以常数，然后再把获得的方程与另一个方程相加，即对方程进行线性组合，那么相应方程中的系数会发生怎样的变化?

### 运算法则

假定某个表达式为一组结构方程的非递归模型，其误差项与所有外生变量不相关，而且所有变量与误差项的均值为 0。同时，m、k、$m_i$、$m_e$ 和 $k_e$ 的定义

也如上一节所述。我们形成如下所述的 m 行、m+k 列的系统矩阵。在系统矩阵的最左侧，依次罗列了以任意顺序排列的模型内生变量，系统矩阵顶端则列出了以任意顺序排列的所有模型变量（包括内生和外生变量）。

系统矩阵的每一行对应定义该行左侧的内生变量的一个结构方程。我们一次只考虑一个结构方程，以确定矩阵中与之对应的输入。在每一列对应方程所包含的每个变量的位置输入一个星号，不论变量在该方程的右边还是左边。另外，在该行的所有其他位置输入 0。[19]对模型中所有其他结构方程（即系统矩阵的其他行）进行同样的操作。结果就形成一个只包含 0 和星号的矩阵，该矩阵提供了使检验模型中任何一个方程可辨识性所需要的全部信息。

在次序条件检验中，需运用运算法则对每个方程单独进行可辨识性的检验。而且，作为第一步，在运算法则中建立次序条件也很有用。因此，对于模型中某个待检验可辨识性的方程，$k_e$ 必须大于或等于 $m_i$ 减去 1。如果 $k_e < m_i - 1$，则该方程必定是不可辨识的，也就无需进行运算法则的其余步骤了。但如果 $k_e \geqslant m_i - 1$，那么接下来的步骤就是必要的。

我们把系统矩阵中对应待检验方程可辨识性的行称为“检验行”。第一步是在检验行上画一条线。[20]同时，在检验行中星号所在的列也画一条线。然后改写系统矩阵：(1)删除画线那一行，即检验行；(2)删除画线的所有的列。剩下的这个子矩阵就是该检验行的收缩矩阵。

我们还必须定义 3 个概念：(1)如果两个矩阵行在矩阵列上的输入完全相同，那么这两个矩阵行就是等同的；(2)对于某矩阵行，从左往右的第一个星号被称做该行的“领头星号”；(3)如果存在领头星号的每个列上的其他输入皆为 0，那么这就是该矩阵的简单形式。[21]为了进一步说明这些定义，以下面的 3×3 矩阵为例：

$$
\begin{array}{cc}
 & \text{列} \\
 & \begin{array}{ccc} 1 & 2 & 3 \end{array} \\
\text{行}\ \begin{array}{c} 1 \\ 2 \\ 3 \end{array} & \begin{bmatrix} * & 0 & * \\ * & * & 0 \\ * & 0 & * \end{bmatrix}
\end{array}
$$

首先，我们看到该矩阵的行 1 和行 3 是等同的，因为这两行在列 2 都是 0，其他列都是星号。其次，这不是一个简单式的矩阵，只要看到列 1 包含了行 1 的领头星号，同时也包含了行 2 的星号，我们就知道这不是简单式。简单式背后的含义将随着讨论的展开而逐渐清晰。从本质上来说，这种定义的策略可以对压缩矩阵进行操作，从而把矩阵转化成简单式。一旦获得了简单式矩阵，我们就能很快判断对应该矩阵中检验行的方程是否可辨识。

分析收缩矩阵的步骤可参见图 3.1。该流程图中的所有步骤是一个完整的检验可辨识性的运算法则。现在，我将举一些例子来说明整个检验模型方程可辨识性的过程。

第一，职业和教育期望模型。我们现在对上文提及的职业期望模型加以扩展，以更接近于 Duncan 等人提出的原初模型(Duncan et al.，1971)。图 3.2 所示的这一模型包含 4 个内生变量：(1)某男性受访青少年的职业期望水平 $X_1$；(2)该受访者朋友的职业期望水平 $X_2$；(3)受访者的教育期望水平 $X_3$；(4)受访者朋友的教育期望水平 $X_4$。我们假定这两个职业期望变量互为因果，两个教育变量之间也是如此。而且，对于受访者及其朋友，我们假定职业期望是教育期望的一个直接原因。我们还假设能获得一个有关智力的指标(只是受访者的智力)，并认为受访者的智力($Z_5$)对他的职业和教育期望都具有直接作用。[22]最后，我们把受访者与朋友的家庭社会经济地位 $Z_6$ 和 $Z_7$ 作为外生变量引入模型中，它们都直接影响职业和教育期望水平。因此，该模型的结构方程组如下所示：

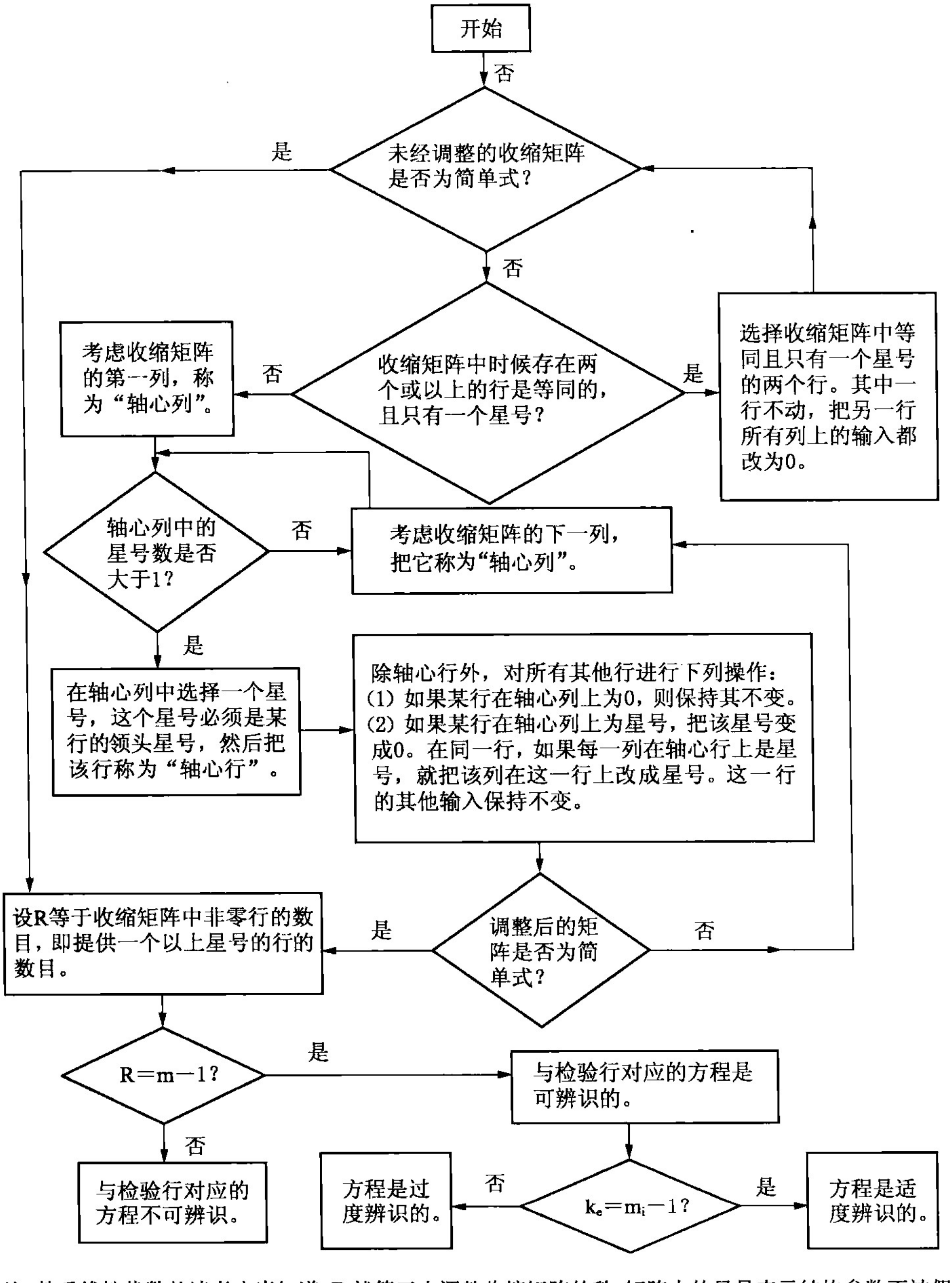
$$X_1 = \beta_{12}X_2 + \gamma_{15}Z_5 + \gamma_{16}Z_6 + \varepsilon_1 \quad [3.6]$$

$$X_2 = \beta_{21}X_1 + \gamma_{27}Z_7 + \varepsilon_2 \quad [3.7]$$

$$X_3 = \beta_{31}X_1 + \beta_{34}X_4 + \gamma_{35}Z_5 + \gamma_{36}Z_6 + \varepsilon_3 \quad [3.8]$$

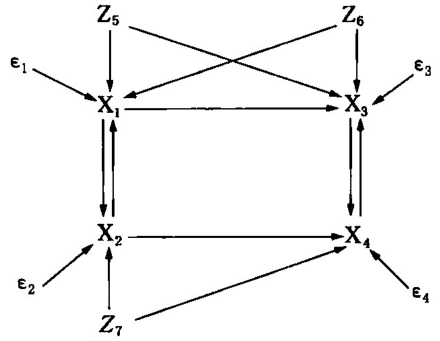
$$X_4 = \beta_{42}X_2 + \beta_{43}X_3 + \gamma_{47}Z_7 + \varepsilon_4 \quad [3.9]$$

假定每个误差项($\varepsilon_1$、$\varepsilon_2$、$\varepsilon_3$ 和 $\varepsilon_4$)与每个外生变量($Z_5$、$Z_6$ 和 $Z_7$)都不相关，而且所有的变量和误差项的均值为 0。对于这一结构方程组来说，$m = 4$，$k = 3$，




注：熟悉线性代数的读者应当知道，R就等于未调整收缩矩阵的秩，矩阵中的星号表示结构参数不被假定为0。

**图3.1 分析收缩矩阵的流程图**




注：假设 $cov(\varepsilon_i, Z_j) = 0$，其中 $i = 1$、2、3 和 4；$j = 5$、6 和 7；$E(X_i) = E(\varepsilon_i) = E(Z_j) = 0$，其中 $i = 1$、2、3 和 4；$j = 5$、6 和 7。$X_1$ 为受访者的职业期望水平，$X_2$ 为朋友的职业期望水平，$X_3$ 为受访者的教育期望水平，$X_4$ 为朋友的教育期望水平，$Z_5$ 为受访者的智力水平，$Z_6$ 为受访者的家庭社会经济地位，$Z_7$ 为朋友的家庭社会经济地位。

**图 3.2　方程 3.6 至方程 3.9 所示的非递归模型的因果图示**

其系统矩阵可表示为如下形式：

$$
\begin{array}{c} \\ X_1 \\ X_2 \\ X_3 \\ X_4 \end{array}
\begin{array}{c}
\begin{array}{ccccccc} X_1 & X_2 & X_3 & X_4 & Z_5 & Z_6 & Z_7 \end{array} \\
\begin{bmatrix}
* & * & 0 & 0 & * & * & 0 \\
* & * & 0 & 0 & 0 & 0 & * \\
* & 0 & * & * & * & * & 0 \\
0 & * & * & * & 0 & 0 & *
\end{bmatrix}
\end{array}
$$

首先来检验定义 $X_1$ 的结构方程 3.6 的可辨识性。对于该方程，$k_e \geqslant m_i - 1$，因为 $k_e = 1$ 且 $m_i = 2$。因此次序条件是满足的，必须继续我们的运算法则。对检验行（行 $X_1$）及列 $X_1$、$X_2$、$Z_5$ 和 $Z_6$（在检验行提供了星号）分别画线，就得到：

$$
\begin{array}{c} \\ X_1 \\ X_2 \\ X_3 \\ X_4 \end{array}
\begin{array}{c}
\begin{array}{ccccccc} X_1 & X_2 & X_3 & X_4 & Z_5 & Z_6 & Z_7 \end{array} \\
\begin{bmatrix}
* & * & 0 & 0 & * & * & 0 \\
* & * & 0 & 0 & 0 & 0 & * \\
* & 0 & * & * & * & * & 0 \\
0 & * & * & * & 0 & 0 & *
\end{bmatrix}
\end{array}
$$

因此,该检验行的收缩矩阵如下:

$$\text{行}\begin{array}{c} \\ 1 \\ 2 \\ 3 \end{array}\begin{array}{c}\text{列} \\ \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \begin{bmatrix} 0 & 0 & * \\ * & * & 0 \\ * & * & * \end{bmatrix}\end{array}$$

接下来用图 3.1 所示的步骤对收缩矩阵进行调整。首先,请注意未经调整的收缩矩阵不是简单式(比如行 2 在列 1 存在领头星号,但列 1 包含了两个星号)。其次,收缩矩阵中不存在两个行是等同的。我们选择列 1 为轴心列(定义参见图 3.1)开始,但由于列 1 存在一个以上的星号(行 2 和行 3),所以我们必须选择行 2 或行 3 为轴心行(定义参见图 3.1)。注意,这一选择并不会影响收缩矩阵的 R 值(定义参见图 3.1),虽然它会影响最终形成的简单式矩阵中 0 和星号的位置。

选择行 2 为轴心行。我们必须逐个检验矩阵中的其他行,即行 1 和行 3。因为行 1 在轴心列(列 1)上是 0,故保持行 1 不变。但行 3 在轴心列上是星号,所以必须进行调整。根据流程图的指示,我们把行 3/列 1 位置的星号改为 0。接着,因为轴心行在列 2 上是星号,所以我们必须在行 3/列 2 的位置上输入星号。当然,在本例中,该位置上已经是星号了,因此这一操作没带来变化。经过这些调整之后,矩阵变成了如下形式:

$$\text{行}\begin{array}{c} \\ 1 \\ 2 \\ 3 \end{array}\begin{array}{c}\text{列} \\ \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \begin{bmatrix} 0 & 0 & * \\ * & * & 0 \\ 0 & * & * \end{bmatrix}\end{array}$$

这还不是简单式矩阵(比如,其中行 3 的领头星号在列 2,但列 2 包含两个星号),因此我们必须重新回到流程图,并以列 2 为轴心列。列 2 包含一个以上的星号,但只有一个是领头星号(在行 3 上),因此必须以行 3 为轴心行。由于行

1在轴心列上为0,故保持行1不变。但行2在轴心列上为星号,必须加以调整。我们把行2/列2的位置变成0,把行2/列3的位置变成星号。经过这些调整,就获得了如下矩阵:

| | | 列 1 | 列 2 | 列 3 |
|---|---|---|---|---|
| 行 | 1 | 0 | 0 | * |
| 行 | 2 | * | 0 | * |
| 行 | 3 | 0 | * | * |

这个矩阵仍然不是简单式,因为列3包含行1的领头星号和其他两个星号。所以我们再次回到流程图,并以列3为轴心列。列3中的3个星号中,只有行1上的是领头星号,所以行1必须成为轴心行。然后我们调整行2和行3,因为它们在轴心列上都是星号。把列3上所有其他星号(即行2/列3和行3/列3)改成0,就得到如下矩阵:

| | | 列 1 | 列 2 | 列 3 |
|---|---|---|---|---|
| 行 | 1 | 0 | 0 | * |
| 行 | 2 | * | 0 | 0 |
| 行 | 3 | 0 | * | 0 |

这是一个简单式矩阵,因为每一列都只有一个星号。所以不需要再对矩阵进行调整。

由于简单式矩阵中所有的行都包含星号,因此 $R=3$,且由于 $m=4$,所以 $R=m-1$。因此,代表 $X_1$ 的结构方程3.6可辨识。最后,由于 $k_e=(m_i-1)=1$,所以可知该方程是适度辨识的。

接下来,检验期望模型中方程3.7的可辨识性。由于 $k_e=2$,$m_i=2$,进而得到 $k_e>m_i-1$,所以该方程的次序条件是满足的。在检验行($X_2$ 所在行)以及 $X_1$、$X_2$ 和 $Z_7$ 所在列上画线,删除所画线上的输入,便得到如下的检验行收缩矩阵:

$$
\begin{array}{cc|cccc|}
 & & \multicolumn{4}{c}{\text{列}} \\
 & & 1 & 2 & 3 & 4 \\
 & 1 & 0 & 0 & * & * \\
\text{行} & 2 & * & * & * & * \\
 & 3 & * & * & 0 & 0
\end{array}
$$

该收缩矩阵不是简单式，而且不存在两个等同的列。以列 1 为轴心列，选择行 3 为轴心行。行 1 在轴心列上是 0，所以保持行 1 不变。行 2 在轴心列上是星号，因此必须进行调整。调整后的矩阵如下：

$$
\begin{array}{cc|cccc|}
 & & \multicolumn{4}{c}{\text{列}} \\
 & & 1 & 2 & 3 & 4 \\
 & 1 & 0 & 0 & * & * \\
\text{行} & 2 & 0 & * & * & * \\
 & 3 & * & * & 0 & 0
\end{array}
$$

然后，列 2 成为轴心列，且要求行 2 成为轴心行。行 1 不变，但行 3 必须进行调整，从而得到下列矩阵，但这仍然不是简单式：

$$
\begin{array}{cc|cccc|}
 & & \multicolumn{4}{c}{\text{列}} \\
 & & 1 & 2 & 3 & 4 \\
 & 1 & 0 & 0 & * & * \\
\text{行} & 2 & 0 & * & * & * \\
 & 3 & * & 0 & * & *
\end{array}
$$

接着，以列 3 为轴心列，行 1 为轴心行。对行 2 和行 3 进行必要的调整，得到如下矩阵：

$$
\begin{array}{cc|cccc|}
 & & \multicolumn{4}{c}{\text{列}} \\
 & & 1 & 2 & 3 & 4 \\
 & 1 & 0 & 0 & * & * \\
\text{行} & 2 & 0 & * & 0 & * \\
 & 3 & * & 0 & 0 & *
\end{array}
$$

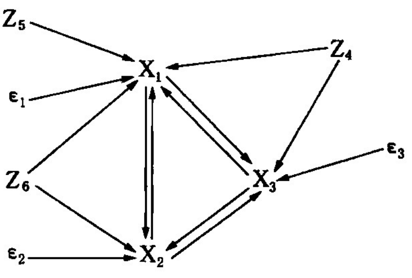
这个矩阵是简单式，因此不需要继续调整。我们已经得到 $R=3$，既然 $m=4$，所以 $R=m-1$。因此，定义 $X_2$ 的结构方程 3.7 可辨识。最后，回顾该方程的次序条件检验，已知 $k_e > m_i - 1$，所以方程 3.7 是过度辨识的。

接下来再看结构方程 3.8，其中 $k_e = 1$，$m_i = 3$。因此方程 3.8 不满足次序条件，所以必然不可辨识。最后，对方程 3.9 进行运算，发现其适度辨识，因为最后结果是 $R = 3 = m - 1$，而且 $k_e = 2 = m_i - 1$。因此，图 3.2 所示的教育和职业期望模型的辨识性如下：结构方程 3.6、结构方程 3.7 和结构方程 3.9 是可辨识的，其中 $X_1$ 和 $X_4$ 的方程是适度辨识的，$X_2$ 的方程是过度辨识的，但定义 $X_3$ 的方程 3.8 是不可辨识的。如果拥有足够多的数据，我们就能合理地估计方程 3.6、方程 3.7 和方程 3.9 中的参数，但无法估计方程 3.8 的参数。因此，我们有必要对模型进行修改以使所有方程都变成可辨识的（将在第 4 章讨论）。

第二，投票行为模型。为了让读者明白该运算法则，我们将对图 1.4 所示的投票模型的扩展模型进行可辨识性检验。我们增加了两个来自 Page 和 Jones（1979）研究的外生变量：(1)父亲的政党认同 $Z_5$，并假定其对个人的政党认同有直接作用；(2)个人的"政党投票史"$Z_6$，假定其对个人政党认同和候选人评价都有直接因果作用。这样，我们的模型就如图 3.3 所示，方程组表达式如下：

$$X_1 = \beta_{12}X_2 + \beta_{13}X_3 + \gamma_{14}Z_4 + \gamma_{15}Z_5 + \gamma_{16}Z_6 + \varepsilon_1 \qquad [3.10]$$

$$X_2 = \beta_{21}X_1 + \beta_{23}X_3 + \gamma_{26}Z_6 + \varepsilon_2 \qquad [3.11]$$




注：假设 $cov(\varepsilon_i, Z_j) = 0$，其中 $i = 1$、2 和 3；$j = 4$、5 和 6；$E(X_i) = E(\varepsilon_i) = E(Z_j) = 0$，其中 $i = 1$、2 和 3；$j = 4$、5 和 6。$X_1$ 为政党认同，$X_2$ 为候选人评价，$X_3$ 为政策/议题立场，$X_4$ 为教育，$Z_5$ 为父亲的政党认同，$Z_6$ 为政党投票史。

**图 3.3 方程 3.10、方程 3.11 和方程 3.12 所示的非递归模型的因果图示**

$$X_3 = \beta_{31}X_1 + \beta_{32}X_2 + \gamma_{34}Z_4 + \varepsilon_3 \quad [3.12]$$

对于该模型，$m = k = 3$，其系统矩阵如下：

$$\begin{array}{c} \\ X_1 \\ X_2 \\ X_3 \end{array} \begin{array}{c} \begin{array}{cccccc} X_1 & X_2 & X_3 & Z_4 & Z_5 & Z_6 \end{array} \\ \begin{bmatrix} * & * & * & * & * & * \\ * & * & * & 0 & 0 & * \\ * & * & * & * & 0 & 0 \end{bmatrix} \end{array}$$

我们很容易就能得出在投票模型中，定义 $X_1$ 的结构方程 3.10 不可辨识，因为通过次序条件检验，$k_e = 0$，$m_i = 3$，进而得到 $k_e < m_i - 1$。但方程 3.11 满足次序条件（$k_e = 2$，$m_i = 3$），因此，删除了系统矩阵中 $X_2$ 所在行以及 $X_1$、$X_2$、$X_3$ 和 $Z_6$ 所在列，得到该方程的收缩矩阵如下：

$$\begin{array}{c} \\ \\ 行 \end{array} \begin{array}{c} 列 \\ \begin{array}{cc} 1 & 2 \end{array} \\ \begin{array}{c} 1 \\ 2 \end{array} \begin{bmatrix} * & * \\ * & 0 \end{bmatrix} \end{array}$$

以列 1 为轴心列，行 2 为轴心行。然后按照规则，对行 1 进行调整，得到一个简单式矩阵如下：

$$\begin{array}{c} \\ \\ 行 \end{array} \begin{array}{c} 列 \\ \begin{array}{cc} 1 & 2 \end{array} \\ \begin{array}{c} 1 \\ 2 \end{array} \begin{bmatrix} 0 & * \\ * & 0 \end{bmatrix} \end{array}$$

因此，R 等于 2，与 $m-1$ 相等，所以该方程是可辨识的。既然 $k_e = m_i - 1$，那么方程 3.11 是适度辨识的。

运算法则同样显示方程 3.12 也是适度辨识的，该方程的收缩矩阵如下：

$$\begin{array}{c} \\ \\ 行 \end{array} \begin{array}{c} 列 \\ \begin{array}{cc} 1 & 2 \end{array} \\ \begin{array}{c} 1 \\ 2 \end{array} \begin{bmatrix} * & * \\ 0 & * \end{bmatrix} \end{array}$$

对方程进行调整后，得到 $R = 2 = m - 1$，而且与前一个方程一样，$k_e = 2 = m_i - 1$。因此，投票模型中，两个方程适度辨识，第三个方程不可辨识，所以也需要修改模型。在讨论如何对不可辨识的模型进行修改之前，我们先回到期望模型这个例子。

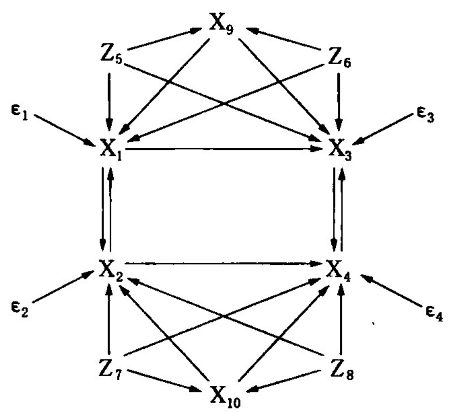
## 次序条件的不充分性：一点说明

在前述两个使用运算法则的例子中，满足次序条件的每个方程最后都被证明是可辨识的。为了抵制认为次序条件是可辨识性的充分且必要条件这一诱惑，我们将以教育和职业期望的扩展模型为例进行说明。这一修正模型如图 3.4 所示，结构方程组表达式如下[23]：

$$X_1 = \beta_{12}X_2 + \beta_{19}X_9 + \gamma_{15}Z_5 + \gamma_{16}Z_6 + \varepsilon_1 \quad [3.13]$$

$$X_2 = \beta_{21}X_1 + \beta_{2,10}X_{10} + \gamma_{27}Z_7 + \gamma_{28}Z_8 + \varepsilon_2 \quad [3.14]$$

$$X_3 = \beta_{31}X_1 + \beta_{34}X_4 + \beta_{39}X_9 + \gamma_{35}Z_5 + \gamma_{36}Z_6 + \varepsilon_3 \quad [3.15]$$




注：假设 $cov(\varepsilon_i, Z_j) = 0$，其中 $i = 1$、2、3、4、9 和 10；$j = 5$、6、7 和 8；$E(X_i) = E(\varepsilon_i) = E(Z_j) = 0$，其中 $i = 1$、2、3、4、9 和 10；$j = 5$、6、7 和 8。$X_1$ 为受访者的职业期望水平，$X_2$ 为朋友的职业期望水平，$X_3$ 为受访者的教育期望水平，$X_4$ 为朋友的教育期望水平，$Z_5$ 为受访者的智力水平，$Z_6$ 为受访者的家庭社会经济地位，$Z_7$ 为朋友的家庭社会经济地位，$Z_8$ 为朋友的智力水平，$X_9$ 为受访者的父母期望水平，$X_{10}$ 为朋友的父母期望水平。

**图 3.4 方程 3.13 至方程 3.18 所示的非递归模型的因果图示**

$$X_4 = \beta_{42}X_2 + \beta_{43}X_3 + \beta_{4,10}X_{10} + \gamma_{47}Z_7 + \gamma_{48}Z_8 + \varepsilon_4 \quad [3.16]$$

$$X_9 = \gamma_{95}Z_5 + \gamma_{96}Z_6 + \varepsilon_9 \quad [3.17]$$

$$X_{10} = \gamma_{10,7}Z_7 + \gamma_{10,8}Z_8 + \varepsilon_{10} \quad [3.18]$$

该模型与图 3.2 所示模型的区别在于，在这里，我们假设能获得有关朋友智商的指标($Z_8$)。该模型还增加了两个内生变量：受访者父母的期望水平($X_9$)和朋友父母的期望水平($X_{10}$)。假定某青少年父母的期望水平影响其教育和职业期望，同时父母的期望水平受该青少年的智力及家庭社会经济地位的影响。

对于这一新模型，使 $m = 6$，$k = 4$，从而形成如下的系统矩阵：

| | $X_1$ | $X_2$ | $X_3$ | $X_4$ | $X_9$ | $X_{10}$ | $Z_5$ | $Z_6$ | $Z_7$ | $Z_8$ |
|---|---|---|---|---|---|---|---|---|---|---|
| $X_1$ | * | * | 0 | 0 | * | 0 | * | * | 0 | 0 |
| $X_2$ | * | * | 0 | 0 | 0 | * | 0 | 0 | * | * |
| $X_3$ | * | 0 | * | * | * | 0 | * | * | 0 | 0 |
| $X_4$ | 0 | * | * | * | 0 | * | 0 | 0 | * | * |
| $X_9$ | 0 | 0 | 0 | 0 | * | 0 | * | * | 0 | 0 |
| $X_{10}$ | 0 | 0 | 0 | 0 | 0 | * | 0 | 0 | * | * |

我们来看模型中的结构方程 3.13。对于该方程，$k_e = 2$，$m_i = 3$，因此 $k_e = m_i - 1$，满足了次序条件。如果按照运算法则继续对方程进行操作，会得到如下收缩矩阵：

| 行 \ 列 | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| 1 | 0 | 0 | * | * | * |
| 2 | * | * | 0 | 0 | 0 |
| 3 | * | * | * | * | * |
| 4 | 0 | 0 | 0 | 0 | 0 |
| 5 | 0 | 0 | * | * | * |

继续进行调整，直至获得简单式矩阵，最后得到 $R = 4$（其实很显然，R 必定小

于 5，因为矩阵中的行 4 在运算法则的任何步骤中都会保持不变）。由于 $m=6$，则 $R \neq m-1$，因此该方程是不可辨识的。我们注意到，即使满足了次序条件，该方程还是不可辨识的。线性组合视角可以说明为什么方程 3.13 是不可辨识的。问题在于，在同一个模型中，方程 3.17 的所有变量（即 $X_9$、$Z_5$ 和 $Z_6$）都在方程 3.13 中出现过。所以方程 3.13 和方程 3.17 的任何非零线性组合所包含的变量都与方程 3.13 完全相同，因此，以线性组合的视角来看，方程 3.13 也是不可辨识的。

总而言之，需要牢记的一点是，次序条件是不充分的，虽然它是表明某些方程不可辨识的有效方法，但却不可用来表明某个方程是可辨识的。如果某个方程满足次序条件，就必须用一个完整的运算流程来检验该方程是否满足秩条件。

# 第 4 章 | 修改不可辨识模型

估计可辨识非递归模型中方程参数的方法有好几种。但是对于不可辨识的方程，没有任何估计方法可以克服可辨识性的缺失，从而产生有意义的参数估计值。那么，这是否意味着面对一个不可辨识模型，我们所能做的只是无奈地耸耸肩，然后转向一个完全不同的研究领域？答案当然是否定的。通过对一个不可辨识模型中的方程增加一些限定，我们可以使其变得"可辨识"。其中的某些限定种类我们已经在第 2 章中讨论过，但至今在社会科学研究中，最为常用的是零限定：假定模型结构方程中的某些变量可以被忽略，即模型中的某些参数等于 0。

确实，在第 2 章中用线性组合视角来检验可辨识性时，我们已经认识到，特定方程中越多的模型变量被排除（即方程中越多的变量参数被限定为 0），该方程就越有可能是可辨识的。我们在上一章中对所列举模型的可辨识性特征的检验与这一经验法则是一致的。比如在图 3.3 所示的投票模型中，唯一不可辨识的方程就是定义 $X_2$ 的方程 3.10。该方程在模型所有方程中是零限定最少的。方程 3.11 和方程 3.12 各自都有两个参数被假定为 0，分别是方程 3.11 中的 $Z_4$ 和 $Z_5$ 以及方程 3.12 中的 $Z_5$ 和 $Z_6$。相反，方程 3.10 中没有零限定，也就是说，模型包含的所有变量都出现在方程 3.10 中。图 3.2 所示的职业和教育期望模型也存在类似情况，唯一一个不可辨识的方程是定义 $X_3$ 的方程 3.8，在模型的 4 个结构方程中，方程 3.8 是删除变量最少的一个。在任何时候，上述经验法则都将是修改不可辨识模型以使其可辨识的中心原则。

借助这一法则，我们继续来看图 3.2 所示的期望模型。我们的基本目标是通过增加零限定，使方程 3.8 变得可辨识。显然，通过删除一个或多个方程中已

有的解释变量来增加零限定是不合适的。假定原初模型是对男性青少年教育和职业期望相关理论的精确界定描述，那么从方程 3.8 中删除某个变量，而在模型的其他方程中仍然保留该变量，就将导致模型的错误设定，即导致模型与构建它的理论脱节。但通过增加影响模型中其他变量（除 $X_3$ 以外）的外生变量，我们或许能增加方程 3.8 的零限定的数目，使其具有可辨识性，但又不会导致模型的错误界定。

比如，当受访者的智力（$Z_5$）作为外生变量被纳入模型时，我们曾假定无法获得衡量受访者的朋友智力的指标。但如果能获得衡量朋友智力的指标，我们就能把它纳入模型，并假定它具有与受访者智力相同的作用机制。也就是说，我们假定朋友的智力（$Z_8$）对 $X_2$（朋友的职业期望水平）和 $X_4$（朋友的教育期望水平）都有直接作用，但对 $X_1$ 或 $X_3$（受访者的职业和教育期望水平）没有直接作用。增加了新的外生变量之后，模型变成了如下形式：

$$X_1 = \beta_{12} X_2 + \gamma_{15} Z_5 + \gamma_{16} Z_6 + \varepsilon_1 \qquad [4.1，与 3.6 相同]$$

$$X_2 = \beta_{21} X_1 + \gamma_{27} Z_7 + \gamma_{28} Z_8 + \varepsilon_2 \qquad [4.2，修改后的 3.7]$$

$$X_3 = \beta_{31} X_1 + \beta_{34} X_4 + \gamma_{35} Z_5 + \gamma_{36} Z_6 + \varepsilon_3 \qquad [4.3，与 3.8 相同]$$

$$X_4 = \beta_{42} X_2 + \beta_{43} X_3 + \gamma_{47} Z_7 + \gamma_{48} Z_8 + \varepsilon_4 \qquad [4.4，修改后的 3.9]$$

在该模型中，与先前相比，$X_3$ 的结构方程有了一个以上的零限定。因此，对期望模型的修改符合有关零限定的经验法则。我们希望这一修改可使模型变得可辨识。的确，运算法则的运用表明，该模型现在变得可辨识了。具体来说，结构方程 4.1 和结构方程 4.2 过度辨识，而方程 4.3 和方程 4.4 则为适度辨识。[24]

接下来，我们尝试一下使图 3.3 所示的投票模型变得可辨识。模型中也只有一个结构方程不可辨识，即决定政党认同的方程 3.10。因此，为了使模型可辨识，我们必须找到可以加入到除方程 3.10 以外的其他模型方程中的外生变量。换言之，我们必须找到至少一个外生变量，确信其对政党认同（$X_1$）没有影响，但却对至少一个其他内生变量具有直接影响。这里我们可以参考 Page 和 Jones（1979）的初始模型。他们提出某人对候选人个人素质的评价影响其对候

选人的评价，但却不直接影响政党认同或政策/议题立场。他们还假定收入对政策/议题立场有直接因果作用，但对政党认同或候选人评价没有直接作用。如果我们把个人素质评价（$Z_7$）和收入（$Z_8$）作为外生变量加入到模型中，我们就得到了如下方程组：

$$X_1 = \beta_{12}X_2 + \beta_{13}X_3 + \gamma_{14}Z_4 + \gamma_{15}Z_5 + \gamma_{16}Z_6 + \varepsilon_1 \qquad [4.5，与 3.10 相同]$$

$$X_2 = \beta_{21}X_1 + \beta_{23}X_3 + \gamma_{26}Z_6 + \gamma_{27}Z_7 + \varepsilon_2 \qquad [4.6，修改后的 3.11]$$

$$X_3 = \beta_{31}X_1 + \beta_{32}X_2 + \gamma_{34}Z_4 + \gamma_{38}Z_8 + \varepsilon_3 \qquad [4.7，修改后的 3.12]$$

如果运用运算法则加以演算，我们将发现，投票模型现在已是可辨识的了。[25]

因此，给不可辨识的多方程模型增加外生变量，可以达到可辨识的目的。当然，人为地给模型增加外生变量对获得模型的可辨识性未必有帮助。我们对上述例子的处理表明，给某个不可辨识的模型增加外生变量需要遵循两类要求：一类是技术性的，一类是理论性的。在技术性层面，为了使模型中某个不可辨识的方程变得可辨识，一个或多个外生变量必须添加到除目标方程以外的其他方程中。但是，并不是模型中新增变量和其他变量的所有因果关系都足以限定模型，从而使目标方程可辨识。而且，也没有简单的规则能指导这些变量之间因果关系的定位。研究者必须尝试对这些因果关系进行不同的组合，并运用运算法则对新形成的方程进行检验。[26]

这些用以克服模型不可辨识性的方法听起来像是游戏，像是我们不断对新的外生变量进行“洗牌”，直至找到合适的为止。但这是一个蹩脚的策略。与运用所有的方法论策略一样，要想这一策略运用得得当和有效，在添加外生变量时，我们必须时刻不忘理论。比如，只有我们有信心认为把个人素质评价变量（$Z_7$）添加到图 3.3 所示的投票模型（新模型的方程表达式是方程 4.5、方程 4.6 和方程 4.7）是从理论出发的，这一操作才合理。换句话说，我们是基于理论认为，$Z_7$ 对政党认同（$X_1$）或政策/议题立场（$X_3$）没有直接因果作用，因此 $Z_7$ 对 $X_1$ 或 $X_3$ 的任何作用都是间接的，是通过 $Z_7$ 对候选人的比较性评价（$X_2$）来发挥作用的。即使这样，我们的理由还不够充分，我们还必须相信，基于理论，$Z_7$ 确实对 $X_2$ 有影响。Klein(1962:18)很好地表述了这一警告：

> 在任何调查中，可辨识性都不能通过简单地增加一些对于系统内部关系而言非常微弱或者边缘性的变量而轻易获得。我们必须增加那些具有实质性意义，却被之前的研究所忽视的变量。

打个比方，假定个人素质评价（$Z_7$）对候选人比较性评价（$X_2$）的影响很微弱，也就是说，$Z_7$ 值在数量上的每一个变化对于 $X_2$ 值数量上的变化影响很小，那么，即使修正后的投票模型是可辨识的，它对于模型真实参数的估计也没有价值。这是因为当非递归模型中新增的外生变量与模型中已有内生变量的相关很微弱时，用来估计参数的统计方法会产生标准误差很大的估计值。这意味着参数的置信区间会变得很宽，这样的经验分析也就不可能使我们用合理的精度对因果作用的大小进行测量。

最后，理论在模型可辨识性中的关键作用表明，理论必须先于研究过程而存在，而非形成于原初模型被证明为不可辨识之后。如果在设计研究项目时以及收集数据之前，研究者没有去寻找能使模型获得可辨识性的外生变量，那么研究者很可能陷入这样的尴尬境地：无法获得理论上合适、技术上也必需的外生变量的数据以使模型可辨识。在这种情况下，要想成功地把一个不可辨识模型转变为可辨识模型，的确希望渺茫。

# 第 5 章 | 估计方法

在本章中，我们将讨论恰当地对非递归模型中的方程参数进行估计的程序。当然，在本章中，我们假定所讨论的模型都是可辨识的，因为不可辨识的模型无法获得有意义的参数估计。在第 1 章中，我们提及对于递归模型中的方程而言，一般最小二乘法回归分析能产生无偏且一致的参数估计值；对于一个可辨识的非递归模型，一般最小二乘法回归并不合适，因为它会产生有偏且不一致的估计值。但幸运的是，我们可以对一般最小二乘法进行修改，使之在运用于非递归模型时，能产生或许有偏但至少一致的参数估计值。

当我们从递归模型转向非递归模型后，一般最小二乘法不再适用的原因在于，我们不能再假定非递归模型中的每个误差项 $\varepsilon_i$ 与结构方程中的所有解释变量包括的误差项不相关。除非一个方程中的误差项与方程所有的解释变量不相关，否则一般最小二乘法产生的参数估计值就会有偏且不一致。为了说明这一点，我们来看一个如下形式的、包含两个变量的回归模型：

$$Y = \beta X + \varepsilon \qquad [5.1]$$

其中，我们假定，X、Y 和 $\varepsilon$ 均值皆为 0，但并不得出误差项 $\varepsilon$ 与解释变量 X 不相关的假定。如果将方程 5.1 乘以 X，则得到：

$$XY = \beta X^2 + X\varepsilon \qquad [5.2]$$

然后对等式两边都取期望值，得到：

$$E(XY) = \beta E(X^2) + E(X\varepsilon) \qquad [5.3]$$

但 $E(XY) = cov(X, Y)$，$E(X\varepsilon) = cov(X, \varepsilon)$，$E(X^2) = var(X)$。[27]因此，方

程 5.3 就变成：

$$\mathrm{cov}(X, Y) = [\beta \mathrm{var}(X)] + \mathrm{cov}(X, \varepsilon) \quad [5.4]$$

进行变换得到：

$$\beta = \frac{\mathrm{cov}(X, Y) - \mathrm{cov}(X, \varepsilon)}{\mathrm{var}(X)} = \frac{\mathrm{cov}(X, Y)}{\mathrm{var}(X)} - \frac{\mathrm{cov}(X, \varepsilon)}{\mathrm{var}(X)} \quad [5.5]$$

这就说明，参数 β 是两个协方差的函数，即 X 和 Y 的协方差以及 X 和 ε 的协方差。如果我们以一般最小二乘法估计值 $\hat{\beta}^{\circ}$ 来估计 β，其中，

$$\hat{\beta}^{\circ} = \frac{S_{YX}}{S_X^2}$$

我们就会得到一个有偏的估计值。从本质上来说，β 的一般最小二乘法估计值只估计了 β 的"cov(X, Y)/var(X)"部分，而忽略了"cov(X, ε)/var(X)"部分。因此，作为 β 的一个估计值，$\hat{\beta}^{\circ}$ 偏离(真实值)的程度取决于方程 5.1 中解释变量与误差项的共变程度。

在本章中，我将介绍两种适用于非递归模型的有限信息估计方法：(1)间接最小二乘法(ILS)；(2)二阶段最小二乘法(2SLS)。这些有限信息方法一次只能估计非递归模型中一个方程的参数，而且只根据这单个方程的限定条件来进行估计。ILS 和 2SLS 确实都是通过对 OLS① 回归这一最常用的有限信息方法的直接修改而形成的。另外，这两种方法如果运用得当，就可以获得非递归模型一致的参数估计值。

除有限信息方法之外，也存在几种适用于非递归模型的完全信息估计方法。这种方法利用所有方程的限定条件，对模型所有方程的参数同时进行估计。完全信息参数估计值的优点在于，它们比有限信息参数估计值更有效。[28] 但另一方面，在社会科学的电脑统计包中，完全信息法并不像有限信息法那么普及，即使能得到，价格也昂贵得多。另外，除了这些实际的限制以外，相比有

① 在下文中，为行文方便，将经常以 OLS 代替"一般最小二乘法"，以 ILS 代替"间接最小二乘法"，以 2SLS 代替"二阶段最小二乘法"。——译者注

限信息估计值,完全信息估计值似乎对模型界定中存在的错误更为敏感,因为由于模型界定错误而导致的一个方程参数估计值的偏差会传递到模型所有其他方程的参数估计值中。正是因为这些原因,完全信息法在社会科学研究中运用很少,所以本书将只讨论有限信息方法。[29]

## 间接最小二乘法和适度辨识的方程

间接最小二乘法是一种适用于非递归模型中的适度辨识方程的参数估计方法。该方法是对用简化式视角来检验模型可辨识性的直接延伸。为了弄清楚这一点,让我们重新来看图 2.4 所示的抗议事件暴力行为模型,该模型在第 2 章中已经用简化式视角进行了检验。在先前的分析中,我们首先把模型的结构方程组(方程 2.20 和方程 2.21)转化成简化式方程 2.25 和方程 2.26。然后,通过说明简化式参数($\pi_{13}$、$\pi_{14}$、$\pi_{23}$和 $\pi_{24}$)能够确定方程 2.20 的结构参数 $\beta_{12}$和 $\gamma_{13}$ 的唯一解,我们证明了结构方程 2.20 是可辨识的。结构方程 2.20 的参数之所以能确定,是因为在假定简化式参数已知的前提下,通过方程 2.27 到方程 2.32 能得出如下解:

$$\beta_{12} = \pi_{14}/\pi_{24} \tag{5.6}$$

$$\gamma_{13} = \pi_{13} - (\pi_{14}\pi_{23}/\pi_{24}) \tag{5.7}$$

当然,在实际情况中,简化式参数不太可能是已知的,因此不能得到真正的结构参数。但我们看到,非递归模型的简化式方程总是可辨识的。进一步来说,根据模型的假定,简化式方程等号右侧的解释变量都是外生变量,而且与方程中的误差项不相关。因此,虽然我们无从得知简化式参数,但我们总能够运用 OLS 回归来获得简化式参数的无偏估计值。以抗议事件暴力行为模型为例,我们能获得简化式参数的 OLS 估计值$\hat{\pi}_{12}$、$\hat{\pi}_{14}$、$\hat{\pi}_{23}$和$\hat{\pi}_{24}$。把这些估计值代入方程 5.6 和方程 5.7,就得到:

$$\hat{\beta}_{12} = \hat{\pi}_{14}/\hat{\pi}_{24} \tag{5.8}$$

$$\hat{\gamma}_{13} = \hat{\pi}_{13} - (\hat{\pi}_{14}\hat{\pi}_{23}/\hat{\pi}_{24}) \tag{5.9}$$

最后，对方程等号右侧进行运算，就能得到结构参数 $\beta_{12}$ 和 $\gamma_{13}$ 的估计值。

上述步骤就是间接最小二乘法，因为它间接运用了一般最小二乘法来获得结构参数的估计值，故而得名。这一运算过程的一般步骤如下：首先，把非递归模型的结构方程转化为简化式；其次，运用 OLS 回归获得简化式参数的无偏估计值；第三，如果该模型中的某个结构方程是适度辨识的，那么就有可能把简化式参数的 OLS 估计值代入与方程 2.27 到方程 2.32 类似的方程中，求出结构方程参数的唯一估计值。

虽然简化式参数的 OLS 估计值是无偏的，但遗憾的是，ILS 最后得到的结构参数估计值并不是无偏的。这是因为结构参数估计值通常是简化式参数估计值的非线性转换，而无偏估计值的非线性转化值通常不是无偏的。对抗议事件暴力行为模型中方程 2.20 的 ILS 估计值就是这样的情况，因为方程 5.8 和方程 5.9 表明，要通过对简化式参数估计值进行乘法或除法运算，才能得到 $\hat{\beta}_{12}$ 和 $\hat{\gamma}_{13}$。虽然 ILS 估计值是有偏的，但它们至少是一致的。所以当样本量接近无穷大时，ILS 估计值就能达到真实值。然而，由于 ILS 估计值仅仅是一致而非无偏的，所以就凸显了大样本对精确估计的重要性，因为我们通常无法保证小样本产生的 ILS 估计值接近总体参数的真实值。

ILS 估计的优点在于其逻辑的简洁性。它是用简化式视角来检验模型可辨识性的自然延伸，因此很容易被理解。但除了最简单的非递归模型之外，对于所有其他实际研究的模型，简化式视角的运算实施非常困难。同样，ILS 在实际运用中也显得过于繁复。虽然 ILS 所需的对简化式方程的 OLS 回归分析在任何电脑统计包中都能找到，但这些统计包往往无法进行最后一步——求得结构参数估计值的运算。幸运的是，另一种估计方法，即二阶段最小二乘法能在一些统计包中找到，它能用于估计非递归模型中一个适度辨识方程的参数。另外，对于一个适度辨识的方程，2SLS 所产生的参数估计值与 ILS 的结果相等。

但前文从简化式视角对可辨识性问题的分析提醒我们，当用于估计一个过度辨识方程的参数时，ILS 会产生一个问题。我们来看图 2.5 所示的职业期望模型中的方程 2.34，该方程是过度辨识的。运用 ILS 的运算步骤，我们发现了

估计 $\beta_{21}$ 的两个不同表达式：

$$\hat{\beta}'_{21} = \hat{\pi}_{23} / \hat{\pi}_{13}$$

$$\hat{\beta}''_{21} = \hat{\pi}_{24} / \hat{\pi}_{14}$$

这确实产生了一个常见的问题：当运用于一个过度辨识的方程时，ILS 会产生方程结构参数估计的多个表达式，并由于抽样误差的存在而导致对同一参数的不同估计值。然而，虽然我们得到了 $\beta_{21}$ 的两个估计值（$\hat{\beta}'_{21}$ 和 $\hat{\beta}''_{21}$），但它们是一致的。因此，当样本量趋近于无穷大时，$\hat{\beta}'_{21}$ 和 $\hat{\beta}''_{21}$ 都接近参数 $\beta_{21}$ 的真实值，因而 $\hat{\beta}'_{21}$ 和 $\hat{\beta}''_{21}$ 之间的差别可以忽略不计。当然，由于不存在无穷大的样本，我们依然必须面对一个过度辨识结构方程存在多个 ILS 估计值问题。显然，如果能够设计一个合理的程序，把这些估计值组合起来，我们应当会获得一个更好的结构参数估计值。从本质上来说，2SLS 就是这样一种方法。

## 二阶段最小二乘法

对 OLS 回归分析的不同修改，产生了另一个适用于非递归模型参数估计的有限信息方法。为了估计结构参数，该方法的第一步同样是估计模型简化式方程的参数。而与 ILS 不同的是，不管对于过度辨识还是适度辨识的方程，2SLS 都能产生唯一的参数估计。对于一个适度辨识的方程来说，2SLS 产生的参数估计与 ILS 产生的完全相同。更进一步来说，如果某非递归模型的所有方程都是适度辨识的，那么 2SLS 和 ILS 产生的估计就与利用完全信息方法产生的估计相同，但后者超出了本书的讨论范围。

为了展开讨论 2SLS 估计方法的步骤，我们将检验图 2.5 所示的职业期望模型中的结构方程 2.34。利用 ILS 来估计这一过度辨识的方程产生了多个参数估计值，我们需要解决这些估计值之间的差别，为每个参数（$\beta_{21}$ 和 $\gamma_{25}$）找出单一的估计值。

2SLS 的出发点是找出 OLS 回归不适合方程 2.34 的原因，而这个原因就是模型并未假定 $X_1$ 与方程的误差项 $\varepsilon_2$ 不相关。有鉴于此，2SLS 就是想办法寻找

一个近似于 $X_1$ 的修正变量，但它与 $\varepsilon_2$ 不相关，这样我们就可以用这个变量来代替方程 2.34 中的 $X_1$，并对修正后的方程使用 OLS。那么，应怎样实践这一想法？正如其名字所提示的，2SLS 包含两个阶段。

在第一阶段，运用 OLS 回归来估计 $X_1$ 的简化式方程 2.37 的参数。这一操作的另一表述就是，$X_1$ 对模型所有外生变量的回归，在这里就是 $Z_3$、$Z_4$ 和 $Z_5$。这样就得到了方程 2.37 的 OLS 简化式参数估计，分别命名为 $\hat{\pi}_{13}$、$\hat{\pi}_{14}$ 和 $\hat{\pi}_{15}$。接着，我们就可以运用这些估计值和下面这个方程：

$$\hat{X}_1 = \hat{\pi}_{13}Z_3 + \hat{\pi}_{14}Z_4 + \hat{\pi}_{15}Z_5 \qquad [5.10]$$

来构建一个新变量 $\hat{X}_1$，称之为"$X_1$ 的工具变量"。$\hat{X}_1$ 有两个重要特性：其一，它"近似"于 $X_1$。实际上，它是通过模型外生变量的线性组合所获得的最近似于 $X_1$ 的一个变量。其二，$\hat{X}_1$ 与方程 2.34 的误差项 $\varepsilon_2$ 不相关。因为模型假定 $Z_3$、$Z_4$、$Z_5$ 与 $\varepsilon_2$ 不相关，而 $X_1$ 只是这 3 个变量的一个线性组合。

在 2SLS 的第二阶段，我们以工具变量 $\hat{X}_1$ 代替方程 2.34 中的 $X_1$，得到：

$$X_2 = \beta_{21}^t\hat{X}_1 + \gamma_{25}^t Z_5 + \varepsilon_2 \qquad [5.11]$$

我们以右上角的"t"来区别本方程与方程 2.34 中的参数。既然已知 $Z_5$ 和 $\hat{X}_1$ 都与 $\varepsilon_2$ 不相关，我们就可以用 OLS 回归来估计方程 5.11 的系数。这样得到的 $\hat{\beta}_{21}^t$ 和 $\hat{\gamma}_{25}^t$ 就是方程 2.34 的结构系数 $\beta_{21}$ 和 $\gamma_{25}$ 的 2SLS 估计值。在某种意义上，$\hat{X}_1$ 作为一个工具，允许我们对原初结构方程进行参数估计。

我们可以把上述程序概括为 2SLS 估计的基本步骤。首先，用 OLS 回归分析估计模型每个简化式方程的参数。在操作层面上，就是对模型的每个内生变量($X_1$，$X_2$，…，$X_m$)分别进行对所有外生变量($Z_{m+1}$，$Z_{m+2}$，…，$Z_{m+k}$)的回归。换言之，就是对模型中每个被假定与误差项不相关的内生变量进行回归。然后，我们用得到的 OLS 参数估计和外生变量的相关数据来建构工具变量 $\hat{X}_1$，$\hat{X}_2$，…，$\hat{X}_m$，每个工具变量都被假定与模型中所有的误差项不相关。我们之所以在第一阶段的回归中将所有的(而非一些)外生变量都作为自变量，是因为我们要让建构出来的工具变量尽量与内生变量近似，而且这些新变量与误差项确实不相关。在第一阶段的回归中，被作为自变量的外生变量数目越多，工具变

量与原初的内生变量就越近似。2SLS 的下一步是，把每一个结构方程中作为解释变量的内生变量 $X_j$ 替换成相应的工具变量 $\hat{X}_j$。替换之后，修正模型中的每个解释变量都可被假定与模型误差项不相关。因此，正如我们在第一步中所做的那样，在 2SLS 的最后一步，我们就可以用 OLS 回归来估计修正后结构方程的参数，所得到的估计就是原初结构方程的 2SLS 估计。

2SLS 可用来估计非递归模型中的任何过度辨识方程的结构参数。而且，它也适用于适度辨识的方程，其估计与 ILS 产生的估计完全相同。因此，估计非递归模型系数的最常用的方法是对模型的所有方程运用 2SLS。

就像 ILS 一样，2SLS 产生的估计值虽可能是有偏的，但却是一致的。不幸的是，我们并不清楚这些估计在小样本中的特性。然而，来自蒙特卡罗模拟研究的一些证据表明，在小样本中，非递归模型参数的 2SLS 估计值比 OLS 估计值的偏差小(Namboodiri et al.，1975:519)。但另一方面，2SLS 估计值似乎比 OLS 的参数估计值有更大的标准误(Namboodiri et al.，1975:519)。2SLS 估计值的标准误的大小部分地取决于第一阶段所建构的工具变量与所替代的内生变量的相似程度。在其他条件都相同的情况下，工具变量与原初内生变量的相关性越高，2SLS 产生的参数估计值的有效性就越高。这再一次提醒我们，为了达到可辨识的目的而增加与内生变量微弱相关的外生变量并无多大用处。从技术上来讲，这样得到的修正模型可以被辨识，但如果 2SLS 所建构的工具变量与模型的内生变量相关度很低，结构参数估计值的标准误差就会很大，从而降低了我们对获得模型真实参数的精确估计值的信心。

如上所述，2SLS 包含两个分别进行的 OLS 回归分析。因此，任何电脑程序或统计包都能进行 2SLS 运算，只要它们能储存回归后因变量的预测值并被用于下一个回归分析中(如 SPSS-X 中的回归程序)。但是，2SLS 估计的运算实际上并不一定要分为两个阶段来做，已经有人设计出了一个线性代数公式可进行一步估计，还有一些常用的统计包(如 SAS 和 BMD)使用该公式来计算 2SLS 估计值。

除了运算方便之外，一步运算确实比两步运算更具优越性。这两种 2SLS 方法都能产生相同且正确的非标准化的结构参数估计值，但两步运算产生的标

准化参数估计却可能是不正确的。更精确地说，两步运算法所产生的外生变量的标准化估计值是正确的，而且与一步运算法所产生的值一样。但两步法所产生的内生变量的标准化参数估计会发生衰减，也就是说，比一步法产生的估计值小。其估计值的衰减是因为 2SLS 第一阶段所产生的工具变量的方差/变异比原初内生变量的方差小。这一减小的方差导致了第二阶段中标准化参数估计的衰减(Kritzer，1976)。

幸好 Kritzer 找到了几种方法来修正这一衰减。其一，可以在进行两步回归之前对模型中的变量进行标准化。经过这一初步的标准化，两步法得到的标准化参数估计就不会发生衰减。其二，可以在两步法演算之后对工具变量参数估计的衰减进行修正，方法是用参数估计值除以第一阶段中用以建构工具变量的多个相关系数(R)，但这些修正衰减的方法只有在需要计算标准化参数估计时才用得到。两步运算法会给出精确的 2SLS 非标准化参数估计，它不需要进行任何修正。当然，Blalock(1967)、Achen(1977)和其他一些研究者认为，只有在极少数的情况下才需要报告标准化的参数估计。因此，在运用两步法来计算内生变量的 2SLS 参数估计时，避免衰减的最好办法就是计算非标准化系数。

但运用一步线性代数计算公式来求 2SLS 的估计值之所以更好，还存在另一个更重要的原因。虽然两步法可以产生精确的非标准化参数估计(以及经过修正的标准化估计)，但却会产生不正确的标准误和多个 $R^2$ 值。在为 2SLS 设计的电脑程序和统计包中，标准误和 $R^2$ 值都是在模型原初解释变量观测值的基础上计算得出的。然而，当用两步法来进行 2SLS 估计时，第一步获得的用于参数估计的标准误和 $R^2$ 值是为第二阶段的 OLS 回归服务的。这些标准误是基于工具变量(而非原初的解释变量)的观测值而计算得出的，因此就不是 2SLS 最终参数估计值正确的标准误。因此，只要有可能，研究者就应该用为 2SLS 设计的电脑程序或统计包来计算 2SLS 参数估计值，这样就能产生正确的标准化和非标准化参数估计以及正确的标准误和 $R^2$ 值。提供 2SLS 程序的最常用的统计软件包是 SAS(Statistical Analysis System)和 TSP(Time Series Processor)，其他软件包还包括 SHZAZM、B345 和 QUAIL。

## 二阶段最小二乘法和多元共线性问题

我们看到,2SLS 适用于非递归模型中适度辨识和过度辨识方程的参数估计。或许有些读者仍然执着地认为,只要采取恰当的统计方法,不可辨识问题就可以被克服,因而他们也许会对一个不可辨识方程采用 2SLS 方法。但这一策略并不奏效。事实上,在有些(并非所有)情况下,由于多元共线性问题,2SLS 对于不可辨识方程将彻底失效。以图 2.4 所示的抗议事件暴力行为模型中的方程 2.21 为例,我们已证明该方程是不可辨识的。如果对方程 2.21 运用 2SLS 方法,我们首先要对方程的内生解释变量 $X_1$ 进行模型外生变量 $Z_3$ 和 $Z_4$ 的回归以产生工具变量 $\hat{X}_1$。到这一步为止,一切都没问题。但在第二阶段,我们用 $\hat{X}_1$ 替代方程 2.21 中的 $X_1$,得到:

$$X_2 = \beta_{21}^{i}\hat{X}_1 + \gamma_{23}^{i}Z_3 + \gamma_{24}^{i}Z_4 + \varepsilon_2 \qquad [5.12]$$

在这里,该方法就出问题了。由于工具变量 $\hat{X}_1$ 是 $Z_3$ 和 $Z_4$ 的线性组合,所以方程 5.12 具有完美的多元共线性特征。也就是说,如果对 $\hat{X}_1$ 进行 $Z_3$ 和 $Z_4$ 的回归,其 $R^2$ 值将等于 1。当存在完美多元共线性时,OLS 回归就不能产生一组唯一的参数估计,只能得到无数组与已知数据相吻合的参数估计。[30]我们再次得到这样一个结论:对于不可辨识的方程,统计手段并不能带来有意义的参数估计。

将 2SLS 运用于可辨识的方程时,也会遇到不那么极端的多元共线性问题。这是因为与原初内生变量相比,第一阶段建构的工具变量与模型外生变量的相关性往往更高。以图 2.5 所示的职业期望模型的参数估计为例,由于 $\hat{X}_1$ 是通过 $X_1$ 对一组外生变量(包括 $Z_5$)的回归建构出来的,所以 $\hat{X}_1$ 与 $Z_5$ 的相关程度肯定要高于 $X_1$ 与 $Z_5$ 的相关性(见方程 5.11)。由于 2SLS 建构的每个工具变量都是模型所有外生变量的一个线性组合,所以一般来说,一个模型方程中包含的被作为解释的变量外生变量的比例越高,其 2SLS 估计产生多元共线性的可能就越大。

当 2SLS 估计出现很高的多元共线性问题时,要发现其存在并不困难。对

检测方法的详细讨论已经超越了本书的范围，但可以在其他文献中找到比较好的处理方法（参见 Lewis-Beck，1980：58—63；Hanushek & Jackson，1977：86—93）。运用 2SLS 时要牢记的一个要点是，仅仅检验结构方程原初解释变量之间的多元共线性是不够的。在很多情况下，虽然原初解释变量之间并不存在显著的多元共线性问题，但在涉及工具变量的第二阶段中，解释变量之间却存在很高的共线性。

要注意的是，即使存在较高的多元共线性，2SLS 产生的参数依然是一致的。较高的多元共线性引发的主要后果是参数估计的标准误较大，因而导致参数估计的置信区间较宽（Hanushek & Jackson，1977：86—90）。克服高多元共线性的唯一理想办法是收集更多的数据，因为样本量的增加会减小参数估计的标准误，从而抵消多元共线性的影响。另一个处理多元共线性常用的策略是删除方程中引发该问题的变量。在 2SLS 中，往往是工具变量导致共线性问题，这就意味着要从结构方程的解释变量中删除一个或更多的内生变量，即等于假定相关的结构参数为 0。这很难说是一种理想的策略，因为这将使获得的模型无法正确界定其背后的理论，而且 2SLS 的估计也将不再一致。

最后，面对高多元共线性问题，有些学者推荐使用脊回归①来估计模型的参数（参见 Hoerl & Kennard，1970；Deegan，1975）。这一方法已经超出了本书的讨论范围。而且，如何在 2SLS 的最后一步运用脊回归而不是 OLS，我们还不清楚。[31] Kritzer（1976：89）对于多元共线性的问题进行了精彩的总结："2SLS 的使用者需要对潜在的多元共线性问题保持敏感，因为一旦出现该问题，可供使用者选择的备用处理方法是有限的。"[32]

## 职业和教育期望模型中的参数估计

为了说明 2SLS，我们将运用该方法来估计方程 4.1 到方程 4.4 的参数。在第 4 章中，我们已知该模型是可辨识的，其中方程 4.3 和方程 4.4 是适度辨识

① 也译为"岭回归"。——译者注

的，而方程 4.1 和方程 4.2 都是过度辨识的。因此，方程 4.3 和方程 4.4 的参数可以用 ILS 或 2SLS 来估计，但方程 4.1 和方程 4.2 只能运用 2SLS 来估计。然而，从计算角度来说，用电脑统计软件包来对所有方程的参数进行 2SLS 估计更简单。用于估计的数据最初是由 Duncan 等人(Duncan et al.，1971)收集的，由对 329 名男性青少年的访谈和测试组成。[33]

对模型进行 2SLS 估计，首先要以模型中的 4 个内生变量($X_1$、$X_2$、$X_3$ 和 $X_4$)对所有的外生变量($Z_5$、$Z_6$、$Z_7$ 和 $Z_8$)进行回归，从而建构 4 个工具变量($\hat{X}_1$、$\hat{X}_2$、$\hat{X}_3$ 和$\hat{X}_4$)，然后用这些工具变量替代方程 4.1 到方程 4.4 中相应的内生解释变量，再用 OLS 回归来估计这些修正方程的参数。所得到的估计就是我们所寻求的 2SLS 估计值。但我们更倾向于使用专为 2SLS 设计的电脑程序，它能计算出 4 个参数估计的准确标准误差。

表 5.1 提供了模型的非标准化 2SLS 参数估计以及它们的标准误。其对方程 4.1 至方程 4.4 也进行了 OLS 参数估计。虽然我们知道 OLS 并不适用于非递归模型，但在表 5.1 中包含这些估计只是为了显示 2SLS 和 OLS 估计的差别(在接下来的讨论中，OLS 估计值以角标“o”来标记，而 2SLS 则以角标“t”标记)。

我们注意到，对于期望模型中的某些参数来说，OLS 估计和 2SLS 估计在数值上区别不大。但对有些参数而言，区别却非常大。其中有 3 个参数的 2SLS 估计值比 OLS 估计值的一半还小：$\hat{\gamma}^{t}_{48} = 0.14$ 而 $\hat{\gamma}^{o}_{48} = 0.50$；$\hat{\beta}^{t}_{31} = 0.19$ 而 $\hat{\beta}^{o}_{31} = 0.46$；$\hat{\gamma}^{t}_{47} = 0.17$ 而 $\hat{\gamma}^{o}_{47} = 0.35$。因此，期望模型的统计结果表明，OLS 回归的结果与 2SLS 所产生的一致的参数估计结果差别很大。我们还注意到一个典型的模式，即 2SLS 参数估计的标准误大于 OLS 参数估计的标准误。对期望模型的所有参数而言，2SLS 估计的标准误的确都大于或等于 OLS 估计的标准误。

如果从 OLS 估计转向更为合适的 2SLS 估计，我们可以得到这样的结论：从一种参数估计技术转向另一种技术时，如果结合不同的参数估计值和不同的标准误来看，研究解释将发生戏剧性的改变。比如，OLS 结果表明，朋友智力水平($Z_8$)对其教育期望水平($X_4$)的影响在统计上具有显著性，因为 $\hat{\gamma}^{o}_{48} = 0.50$，

**表 5.1 职业和教育期望模型方程 4.1 至方程 4.4 的参数估计**

| 解释变量 | 结构参数 | 非标准化参数估计值（括号中为标准误） | |
|---|---|---|---|
| | | 2SLS | OLS |
| 方程 4.1 | | | |
| $X_2$ | $\beta_{12}$ | 0.40(0.10) | 0.30(0.05) |
| $Z_5$ | $\gamma_{15}$ | 0.65(0.12) | 0.70(0.12) |
| $Z_6$ | $\gamma_{16}$ | 0.35(0.12) | 0.41(0.11) |
| 方程 4.2 | | | |
| $X_1$ | $\beta_{21}$ | 0.34(0.12) | 0.26(0.05) |
| $Z_7$ | $\gamma_{27}$ | 0.35(0.12) | 0.39(0.11) |
| $Z_8$ | $\gamma_{28}$ | 0.82(0.13) | 0.87(0.11) |
| 方程 4.3 | | | |
| $X_1$ | $\beta_{31}$ | 0.19(0.70) | 0.46(0.05) |
| $X_4$ | $\beta_{34}$ | 0.20(0.30) | 0.12(0.05) |
| $Z_5$ | $\gamma_{35}$ | 0.50(0.26) | 0.32(0.11) |
| $Z_6$ | $\gamma_{36}$ | 0.53(0.26) | 0.43(0.10) |
| 方程 4.4 | | | |
| $X_2$ | $\beta_{42}$ | 0.77(1.23) | 0.42(0.05) |
| $X_3$ | $\beta_{43}$ | 0.11(0.40) | 0.12(0.04) |
| $Z_7$ | $\gamma_{47}$ | 0.17(0.51) | 0.35(0.09) |
| $Z_8$ | $\gamma_{48}$ | 0.14(1.04) | 0.50(0.10) |

是标准误(0.10)的 5 倍。相反，2SLS 结果显示，朋友智力水平对其教育期望水平的影响 ($\hat{\gamma}'_{48} = 0.14$) 在规定的显著性水平上并不存在统计上的意义，实际上，2SLS 的估计值还不到其标准误(1.04)的 20%。

期望模型中某些方程较大的标准误值使我们有理由怀疑多元共线性问题的存在。最常用的检测多元共线性的方法是观察方程中所有自变量之间的二元相关系数。虽然高多元共线性总能表现在自变量之间的二元相关系数上，但一般来说，高共线性并不必然伴随着较高的二元相关系数。更好地检测方程中多元共线性的办法是检验一组多元相关，尤其是每个自变量对方程中所有其他自变量的回归方程的决定系数 $R^2$ 值(参见 Lemieux, 1978)。我们可以用 $R_j^2$ 来标记这些多元相关，即自变量($X_j$ 或 $Z_j$)对方程中所有其他自变量的回归方程的

决定系数 $R^2$，而用$\hat{R}_j^2$ 来标记在 2SLS 的第二阶段回归中，工具变量$\hat{X}_j$ 对所有其他自变量的回归方程的决定系数 $R^2$。

这一检测方法表明，当运用 2SLS 对方程 4.1 和方程 4.2 进行参数估计时，多元共线性并不表现为一个极端的数值。在方程 4.1 的第二阶段回归中，最大的 $R_j^2$ 是 $\hat{R}_2^2 = 0.37$；对于方程 4.2，最大的 $R_j^2$ 是 $\hat{R}_1^2 = 0.51$。另一方面，方程 4.3 和方程 4.4 参数估计的第二阶段回归则存在很高的多元共线性。对于方程 4.3，在 2SLS 的第二阶段回归中，当我们以工具变量 $X_1$ 或 $X_4$ 对解释变量进行回归时，得到的决定系数 $R^2$ 值($\hat{R}_1^2$ 和$\hat{R}_4^2$)超过了 0.98。较高的多元共线性主要是由于 2SLS 第二阶段方程中的工具变量($\hat{X}_1$ 和$\hat{X}_4$)之间的高相关，$r_{\hat{X}_1\hat{X}_4} = 0.87$。同样，在方程 4.4 的第二阶段回归中，以工具变量($\hat{X}_2$ 和$\hat{X}_3$)对解释变量进行回归，也得到了大于 0.96 的 $R^2$ 值($\hat{R}_2^2$ 和$\hat{R}_3^2$)。

有意思的是，在上述两个方程 2SLS 第二阶段的回归中，充分展示了多元共线性问题并不存在于原初方程的解释变量之间。比如，与方程 4.3 参数估计的第二阶段回归中得到的$\hat{R}_4^2$ 为 0.99 不同，以原初变量 $X_4$ 对方程 4.3 中的其他变量进行回归，所得到的 $R^2$ 值仅为 0.17。这就用实例证明了在运用 2SLS 时，仅对结构方程的原初解释变量进行多元共线性检验是不够的。即使在用工具变量替代原初变量之后出现高共线性问题时，原初变量间的多元共线性仍可能很低。

即使已经检测到 2SLS 估计存在多元共线性问题，在这个阶段的分析中，我们也很难解决这一问题。解决多元共线性最好的办法就是通过扩大样本量来增加所能获得的信息量。一个拥有 329 个个案的样本是一个相当小的样本，实质性地增加样本量可以减轻多元共线性问题，使产生的 2SLS 参数估计的标准误变小。当然，对一个既定样本数据的操作，使我们不可能增加其样本量。另一个可行的方法是从某些结构方程中删除某些解释变量。无疑，从方程 4.3 中删除 $X_1$ 或 $X_4$，或从方程 4.4 中删除 $X_2$ 或 $X_3$，都能极大地降低 2SLS 估计中的多元共线性问题。但是，多元共线性的降低是以削弱对期望模型的正确界定为代价的。删除这些变量需要下列假定中的一个：(1)职业期望不是教育期望的原因；(2)某人的教育期望并不受其同辈群体的影响。而这两个假定都与经验

分析所含的理论相违背。

考虑到这些限制，我们不得不接受这一事实，即现有的数据并不足以获得我们确信能获得的方程 4.3 和方程 4.4 的参数估计。2SLS 第二阶段回归中存在的严重多元共线性问题导致所获得的参数估计会在不同样本之间大幅度波动。另一方面，在方程 4.1 和方程 4.2 的参数估计中，并未出现较高的多元共线性。因此，尽管已有数据存在局限性，我们依然能够获得职业期望模型中变量效应的有效估计值。也就是说，我们有足够的证据来支持下列假设：某个男性青少年的职业期望与其同伴的职业期望互为因果。

## 抗议事件暴力行为修正模型

作为运用 2SLS 的第二个例子，我们将对图 2.4 所示模型进行扩展，用 Kritzer(1977)的原初模型来研究抗议事件中暴力行为的发生。该模型如图 5.1 所示，方程组表达式如下：

$$X_1 = \beta_{12}X_2 + \beta_{15}X_5 + \gamma_{19}Z_9 + \varepsilon_1 \quad [5.13]$$

$$X_2 = \beta_{21}X_1 + \gamma_{23}Z_3 + \gamma_{24}Z_4 + \gamma_{26}Z_6 + \gamma_{27}Z_7 + \gamma_{28}Z_8 + \varepsilon_2 \quad [5.14]$$

$$X_5 = \beta_{51}X_1 + \gamma_{53}Z_3 + \varepsilon_5 \quad [5.15]$$

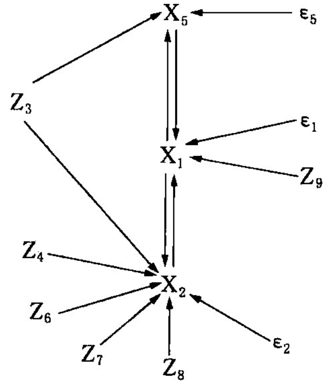



图 5.1 方程 5.13 至方程 5.15 所示的非递归模型的因果图示

$X_1$、$X_2$、$Z_3$ 和 $Z_4$ 如图 2.4 模型中的定义。但我们增加了一个内生变量 $X_5$ 以表示抗议事件中的逮捕情况。Kritzer 假设逮捕情况是影响抗议者暴力水平($X_1$)的一个因素,但反过来,抗议者暴力水平也影响逮捕情况。另外,我们还增加了一些外生变量,在此有必要对其中一些加以说明。3 个外生变量被认为对 $X_2$(警察的暴力水平)有直接作用:抗议者下流手势的使用情况($Z_6$)、抗议者脏话的使用情况($Z_7$)以及现场警察的种类和数量($Z_8$)。最后,我们还加入了一个外生变量 $Z_9$(抗议者对非暴力的执行程度),它被假定对 $X_1$ 有直接作用。

秩条件运算法结果表明,抗议事件暴力行为修正模型的方程 5.14 是适度辨识的,而方程 5.13 和方程 5.15 都是过度辨识的。[34] 因此,如果获得恰当的数据,我们就能运用 2SLS 获得模型参数的有意义的估计值。我利用的数据是 Kritzer 收集的,包含对 126 个抗议活动的观察。[35]

为了便于说明,我计算了标准化的参数估计,以表明 2SLS 估计既可以用非标准化形式,也可以用标准化形式。表 5.2 提供了 2SLS 估计和 OLS 估计。[36]

**表 5.2 抗议事件暴力行为模型方程 5.13 至方程 5.15 的参数估计**

| 解释变量 | 结构参数 | 标准化参数估计值(括号中为标准误) | |
|---|---|---|---|
| | | 2SLS | OLS |
| 方程 5.13 | | | |
| $X_2$ | $\beta_{12}$ | 1.54(0.51) | 0.36(0.07) |
| $X_5$ | $\beta_{15}$ | −0.45(0.43) | 0.39(0.07) |
| $Z_9$ | $\gamma_{19}$ | −0.05(0.13) | −0.18(0.07) |
| 方程 5.14 | | | |
| $X_1$ | $\beta_{21}$ | 0.75(0.27) | 0.41(0.08) |
| $Z_3$ | $\gamma_{23}$ | 0.10(0.08) | 0.13(0.07) |
| $Z_4$ | $\gamma_{24}$ | 0.08(0.10) | 0.14(0.08) |
| $Z_6$ | $\gamma_{26}$ | −0.01(0.13) | 0.11(0.09) |
| $Z_7$ | $\gamma_{27}$ | −0.02(0.09) | 0.03(0.08) |
| $Z_8$ | $\gamma_{28}$ | 0.02(0.10) | 0.09(0.08) |
| 方程 5.15 | | | |
| $X_1$ | $\beta_{51}$ | 0.68(0.10) | 0.54(0.06) |
| $Z_3$ | $\gamma_{53}$ | 0.25(0.07) | 0.29(0.07) |

与先前的例子一样，一些 OLS 参数估计与更加适合的 2SLS 估计差别很大。差别最大的是对 $\beta_{15}$ 的估计，2SLS 的估计是 −0.45，而 OLS 的估计却恰恰相反，是 0.39。[37] 但 $\beta_{12}$ 的两个估计值也存在本质差别，其中 $\hat{\beta}^{t}_{12} = 1.54$，而 $\hat{\beta}^{o}_{12} = 0.36$。同样，2SLS 估计的标准误也依然大于 OLS 估计的标准误①，对于模型中的所有参数，2SLS 估计的标准误都大于或等于 OLS 估计的标准误。

对 $R^2_j$ 系数的检验表明，方程 5.15 中的共线性并没有严重到威胁参数的估计。由于在该方程中只有两个解释变量，第二阶段任一解释变量的 $R^2_j$ 系数与 $\hat{X}_1$ 和 $Z_3$ 之间的二元相关系数的 $r^2$ 相等，为 0.16。

相反，较高的共线性问题则存在于方程 5.13 和方程 5.14 参数估计的第二阶段回归中。对于方程 5.14，$\hat{R}^2_1 = 0.85$；对于方程 5.13，$\hat{R}^2_2 = 0.87$，$\hat{R}^2_5 = 0.85$。虽然这些 $R^2_j$ 值并没有大到期望模型中某些方程所显示的 0.98 甚或更大，但其大小却足以引起我们的关注。最好的解决办法还是寻找一个样本量大于 126 的数据。次优的选择是删除模型中一个或多个内生变量。但要注意到，这在减弱多元共线性问题的同时，增加了模型界定错误，因而增加了 2SLS 估计中的额外误差。Kritzer 权衡了利弊之后，决定从方程 5.13 中删除变量 $X_5$ 并接受模型界定错误所带来的后果。删除 $X_5$ 后，$\hat{\beta}^{t}_{12}$ 从 1.54 降到了 1.05，而 $\hat{\gamma}^{t}_{19}$ 在数值上有所增加，由 −0.05 变为 −0.08。$\hat{\beta}^{t}_{12}$ 在数值上发生了较大的变化。确实，当从一个存在高共线性的方程中删除某个解释变量后，参数估计值发生实质性变化的情况并不少见(Lewis-Beck，1980：60)。

与 Kritzer 的选择不同，我不倾向于通过删除变量来解决多元共线性问题。我相信删除变量所导致的模型界定错误的后果往往比 2SLS 中存在的多元共线性问题更严重。我认为一个准确界定的模型更重要，其次才是对多元共线性进行检测，然后清晰地报告共线性的程度，而非有意去估计一个界定错误的模型。如果某些方程存在的多元共线性问题很严重，那么研究者在解释这些方程的估计系数的意义时就要格外小心，因为这些估计值在不同的样本之间存在大幅度

① 原书该句为“the pattern of larger standard errors for OLS estimates than 2SLS estimates continues to hold”，参照上下文，当为印刷错误，故更正。——译者注

的波动。幸运的是,这些方程参数估计的标准误的大小以及相关的显著性检验提供的信息告诉了我们在解释特定模型时需要谨慎的程度。从根本上来说,多元共线性是一个缺乏足够数据的问题。当数据不足又不能进一步获得数据时,我们必须接受已有数据会产生具有较大标准误的参数估计的事实,并在认识到其局限的同时,竭尽全力地从参数估计中搜集信息。

# 第 6 章 | 结论

我们已经看到，虽然所有的递归模型都是可辨识的，但是递归模型所要求的严格假定（即模型是分层的、变量之间不存在互为因果的关系以及误差项之间互不相关）在很多社会科学研究中是不现实的。当递归模型的假定不能被满足时，我们就必须使用非递归模型。非递归模型不能保证其可辨识性，而且即使这类模型是可辨识的，适用于递归模型参数估计的 OLS 回归也不适用于非递归模型。对于非递归模型，OLS 参数估计值不仅有偏，而且不一致。值得庆幸的是，只要遵循模型误差项与所有外生变量不相关这一假定，很多非递归模型还是可以通过零限定来达到可辨识的。进一步来说，当某个非递归模型可辨识，而且拥有足够的数据，那么就能选择 OLS 回归以外的方法来产生一致的模型参数估计，2SLS 就是其中最常用的一种。

对大多数希望使用非递归模型的社会科学家来说，本书应当为所讨论的主题提供丰富的背景知识。但是，由于篇幅所限，本书无法一一介绍其他与非递归模型有关并应该引起重视的主题，而仅在全书即将结尾之际，罗列其中一些一般性的话题。

## 其他方程形式

本书主要关注包含结构方程组形式的非递归模型，它们本质上都是线性和累加的。但就像回归模型可以加以修改，从而允许非线性（比如多项式和指数模型）和非累加或者相互作用的模型界定（比如乘法模型），非递归模型中的结构方程组也可以如此。对这些形式的讨论可参见 Tufte（1974：第 3 章）

以及 Hanushek 和 Jackson(1977:96—108)。

## 滞后内生变量

一些被准确界定的社会科学理论会产生这样一种模型，其结构方程组中的一些解释变量是方程因变量的先前值，即滞后内生变量。举个例子，一些预算政治学模型把政府前一年的拨款作为一个影响当前拨款的解释变量。对本书所讨论的可辨识性检测和参数估计的方法稍作调整后，就能兼顾滞后内生变量。有关这一话题的简要讨论可见本书附录 1。

## 其他类型的可辨识性限定

第 3 章讨论的运算法则和第 4 章讨论的把不可辨识的非递归模型修改成可辨识的模型的方法，在应用时存在局限性，它们用来获得模型可辨识性的限定仅有以下两种：(1)零限定，即假定模型中某些参数为 0；(2)假定模型中每个误差项都与所有外生变量不相关。虽然这两个限定在社会科学中最常用于辨识非递归模型，但还存在很多其他的限定可以使用。它们包括：(1)假定模型中的一对参数存在已知的比例关系；(2)已知模型中一些误差项的方差，或已知一对或多对误差项之间方差的比例关系；(3)假定模型中一对或多对误差项之间互不相关。附录 2 简要讨论了组群递归模型，该模型假定模型中某些误差项互不相关。有关辨识限定的更为详尽的讨论，请参见 Fisher(1966)。

## 其他的估计方法

本书只提供了一种最适用于非递归模型的有限信息估计方法。另外，还存在几种完全信息估计方法。正如第 5 章提及的，这些完全信息方法通常能产生比有限信息方法更有效的估计值。因此，当多元共线性导致 2SLS 的参数估计存在较大标准误时，这些方法就显得格外有用。Christ(1966)对这几种完全信

息方法进行了详细的讨论。

## 不可观测变量

本书假定非递归模型中的变量能够被一个样本观测到，而且没有测量误差。在很多情况下，这一假设是无法实现的。比如，研究者或许想检验一个理论，而该理论包含无法观测到的变量。以 Duncan 等人(Duncan et al. ，1977)认定已是相当完善的教育和职业期望模型为例，其中一个变量是青少年的抱负水平，但在当时的条件下，研究者是无法测量这个变量的。

在其他情况下，研究者或许能够测量一个理论中的所有变量，但其中一些很可能存在测量误差。这是因为社会科学理论中的很多变量都是抽象的概念，不能直接被观测到(比如，异化、社会经济地位和智力)。因此，当为了检验理论而测量这些变量时，我们被迫依赖这些概念的可观测指标。如果我们仅仅把这些指标代入多方程模型的方程之中，变量就会存在测量误差，这就打破了模型的假设，根据测量误差的不同类别，参数估计就可能有偏或不一致。一个更恰当的模型界定应该既包括不可观测的抽象概念，也包括可观测的指标，并且清晰地表述抽象概念之间、抽象概念和可观测指标之间的因果联系。

当一个多方程模型包含不可观测变量时，本书所讨论的方法则既不能用于检验其可辨识性，又不能用于估计其参数。但在某些情况下，这类模型是可辨识的[38]，此时，LISREL 技术就可以用来进行参数估计。从某种意义上说，LISREL 是多指标模型的一个强有力的延伸，多指标模型是 Sullivan 和 Feldman(1979)著作的主题。有关 LISREL 的精彩讨论(但需要读者具备线性代数知识)，可参见 Goldberger 和 Duncan(1973)编写的一本书(尤其是其中 Jöreskog 的文章)以及 Long(1983)的专著。

# 附 录

## 附录1：非递归模型中的滞后内生变量

在第1章中，我提到在本书中，“外生变量”这一概念被等同于前定变量来使用，因为社会科学中绝大多数的前定变量都是外生的。然而，在某些情况下，我们需要发展一个把滞后内生变量作为前定变量的模型。在这种情况下，本书所有关于外生变量的提法同样适用于滞后内生变量。值得一提的是，在运用第3章的运算法则时，滞后内生变量可以被当做外生变量来处理。

但是，并不是在所有的情况下滞后内生变量都可以被当做前定变量来处理的。比如，有一个包含一个结构方程的非递归模型，其中内生变量 $X_{i(t)}$ 作为因变量放在方程的右边，并且解释变量中包含滞后内生变量 $X_{i(t-1)}$：

$$X_{i(t)} = \gamma_{t,t-1}X_{i(t-1)} + (\sum_{j=2}^{k}\gamma_{ij}Z_j) + (\sum_{j=1}^{m}\beta_{ij}X_j) + \varepsilon_{i(t)} \qquad [A1.1]$$

为了把 $X_{i(t-1)}$ 恰当地处理成前定变量，我们必须假定方程 A1.1 中的误差项 $\varepsilon_{i(t)}$ 与 $X_{i(t-1)}$ 不相关。那么，这个假定意味着什么？

为了回答这一问题，我们来考虑把 $X_{i(t-1)}$ 作为因变量的这一“隐含”结构方程。假定产生 $X_{i(t-1)}$ 的流程与产生 $X_{i(t)}$ 的流程完全相同，该隐含结构方程就是：

$$X_{i(t-1)} = \gamma_{t-1,t-2}X_{i(t-2)} + (\sum_{j=2}^{k}\gamma_{ij}Z_j) + (\sum_{j=1}^{m}\beta_{ij}X_j) + \varepsilon_{i(t-1)} \qquad [A1.2]$$

其中，误差项 $\varepsilon_{i(t-1)}$ 被假定反映了那些影响 $X_{i(t-1)}$ 但未被明确纳入模型的变量。我认为，$\varepsilon_{i(t)}$ 与 $X_{i(t-1)}$ 不相关的假设等同于 $\varepsilon_{i(t)}$ 和 $\varepsilon_{i(t-1)}$ 不相关的假设，前一个假

设是我们把 $X_{i(t-1)}$ 处理成前定变量所必需的，后一个假设常被认为缺乏自动相关(参见 Ostrom, 1980)。为了证明这一等价关系，我们注意到，如果 $X_{i(t-1)}$ 和 $\varepsilon_{i(t)}$ 不相关，那么 $\varepsilon_{i(t)}$ 和 $\varepsilon_{i(t-1)}$ 也必然不相关，因为 $\varepsilon_{i(t)}$ 和 $\varepsilon_{i(t-1)}$ 之间非零相关，加上 $\varepsilon_{i(t-1)}$ 是 $X_{i(t-1)}$ 原因的假定，表明了 $\varepsilon_{i(t)}$ 和 $X_{i(t-1)}$ 之间非零的相关。另外，$\varepsilon_{i(t)}$ 和 $\varepsilon_{i(t-1)}$ 不相关也确保了 $\varepsilon_{i(t)}$ 和 $X_{i(t-1)}$ 之间缺乏相关①。

因此，我们看到，为了使非递归模型中的一个滞后内生变量能够被合理地当做前定变量来处理，我们必须假定该内生变量的误差项不存在自相关，而且该变量的滞后值被纳入模型中。这等于假定由影响 $X_{i(t)}$ 的误差项 $\varepsilon_{i(t)}$ 组成的、变量与在前一时间段 $t-1$ 上影响 $X_i$ 值的那些因素完全不相关。很显然，在很多实际情况中，这一假设是不合理的。如果这一假设是不合理的，那么将模型中的滞后内生变量当做前定变量来处理就是不合适的。

当无法假定不存在自相关时，把模型中的一个滞后内生变量处理成内生变量就更为合理。当然，一般来说，把变量作为内生而非前定的变量会增加模型不可辨识的可能性。更进一步来说，即使模型是可辨识的，把滞后内生变量处理成严格的内生变量意味着，当运用 2SLS 来估计结构参数时，我们就不得不为这一滞后内生变量建构一个工具变量。这样做的后果是，在第二阶段会引发高共线性问题的几率大大增加。尽管把滞后内生变量处理成内生的而非前定变量的复杂性有所增加，但也不可以仅仅为了方便而把滞后内生变量处理成前定变量。只有研究者确信不存在自相关时，才能把滞后内生变量处理成前定变量。

## 附录 2：组群递归模型

在本书中，我把注意力仅仅放在遵循下列有关误差项假设的非递归模型

---

① 原文为"it can be shown that a *nonzero* correlation between $\varepsilon_{i(t)}$ and $\varepsilon_{i(t-1)}$ ensures a lack of correlation between $\varepsilon_{i(t)}$ and $X_{i(t-1)}$"，根据上文，此句应该是"it can be shown that a *zero* correlation between ..."。——译者注

上：(1)对于所有 i 和 j，都有 $cov(\varepsilon_i, Z_j) = 0$，即模型中的每个误差项与所有外生变量都不相关；(2)对于所有 i，都有 $E(\varepsilon_i) = 0$，即所有误差项的均值都为 0。如果研究者想对误差项进行一组更强的限定，即进一步假定某些误差项之间互不相关，那么第 3 章用来检验可辨识性的运算法则就不再适用，除非对其进行修改。

对于假定误差项互不相关的模型，一般来说，该运算法则已不再适用，但对其进行修改后，该法则还能处理一组通常被称为“组群递归”的模型。从某种意义上来说，组群递归模型兼具递归和非递归模型的特征。具体来说，组群递归模型中的某些结构方程能被归为一个“组群”(或子类)，当同一组群内的变量互为因果关系时，组群之间也就存在严格的递归关系(单向因果关系和误差项互不相关)。

来看图 A2 所示的非递归模型，其方程表达式如下：

$$X_1 = \beta_{12} X_2 + \gamma_{15} Z_5 + \varepsilon_1 \qquad [A2.1]$$

$$X_2 = \beta_{21} X_1 + \gamma_{26} Z_6 + \varepsilon_2 \qquad [A2.2]$$

$$X_3 = \beta_{31} X_1 + \beta_{34} X_4 + \gamma_{35} Z_5 + \varepsilon_3 \qquad [A2.3]$$

$$X_4 = \beta_{42} X_2 + \beta_{43} X_3 + \gamma_{46} Z_6 + \varepsilon_4 \qquad [A2.4]$$

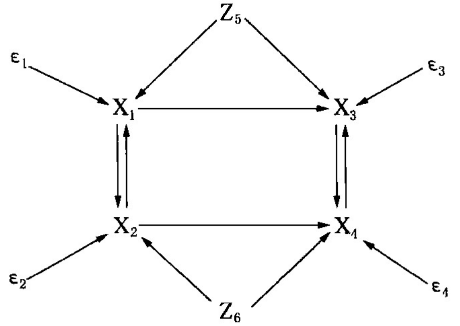



**图 A2　方程 A2.1 至方程 A2.4 所示的非递归模型的因果图示**

如果仅给定本书正文的假设，即对于所有的 i 和 j，都有 $E(\varepsilon_i Z_j) = 0$，对于所有的 i，都有 $E(X_i) = E(Z_j) = E(\varepsilon_i) = 0$，那么读者就可以用运算法则证明方程 A2.1

和 A2.2 是适度辨识的(在这两个方程中，$R = 3$ 且 $m = 4$)，而方程 A2.3 和 A2.4 则是不可辨识的(在这两个方程中，$k_e < m_i - 1$)。但是,如果我们准备对模型中的误差项进行更强的假设,那么该模型就可以从概念化的角度被视为组群递归模型,从而可辨识。具体而言,如果我们额外假定 $\varepsilon_3$ 与 $\varepsilon_1$、$\varepsilon_2$ 都不相关，$\varepsilon_4$ 与 $\varepsilon_1$、$\varepsilon_2$ 都不相关,那么模型就变成了组群递归,其中结构方程 A2.1 和结构方程 A2.2,结构方程 A2.3 和结构方程 A2.4 分别组成了两个组群。在额外的假设之下，$X_1 - X_2$ 组群和 $X_3 - X_4$ 组群内部仍然存在互为因果关系,但因为 $X_1$ 对于 $X_3$ 以及 $X_2$ 对于 $X_4$ 来说是完全前定的,所以两个组群之间是递归关系。

本书中的运算法则经过修改之后就适用于组群递归模型。完全模型中的每一个组群都能被当做一个单独的模型来处理,并对该组群内部的方程实施运算法则而忽略所有组群外的方程。但是当依靠外生和内生变量来计算特定组群中方程的 m、k、$m_i$ 和 $k_e$ 时,任何出现在组群内但是却由组群外的结构方程决定的内生变量(即任何没有出现在组群的内部方程左侧的内生变量)必须被当做外生变量来对待。这样,整个组群递归模型可辨识性的充分条件就是,运算法则显示模型每个组群内的所有方程在组群内部都是可辨识的。

运用这一修改后的运算法则,可以证明图 A2 所示的组群递归模型是可辨识的。具体来说,对于 $X_1 - X_2$ 组群中的两个方程(即方程 A2.1 和方程 A2.2)，$m = k = 2$，$R = 1$，所以 $R = m - 1$，因此两个方程在 $X_1 - X_2$ 组群内部都是可辨识的。对于 $X_3 - X_4$ 组群中的两个方程(即方程 A2.3 和方程 A2.4)，$k = 4$ (因为 $X_1$ 和 $X_2$ 与 $Z_5$ 和 $Z_6$ 一起被当做外生变量)，$m = 2$($X_3$ 和 $X_4$ 被当做内生变量)，$R = 1$，所以对于两个方程,都有 $R = m - 1$，因此方程 A2.3 和方程 A2.4 在组群内部也是可辨识的。

我们再次看到了有关误差项的假设在决定一个模型是否可辨识方面的重要性。如果仅给出本书正文的那些假设,图 A2 所示的模型就是不可辨识的,但是如果设定了组群递归模型所需的假设,该模型就变成可辨识的了。因此,当我们假设模型中误差项不相关时,必须格外谨慎,除非在模型的现实情境中具有合理性,否则我们就不该作出这样的假设。

## 附录3：“期望”模型方程4.1至方程4.4的参数估计中使用的所有变量的标准差和二元积矩相关系数矩阵

| | 原初变量+ | | | | | | | | 工具变量+ | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $X_1$ | $X_2$ | $X_3$ | $X_4$ | $Z_5$ | $Z_6$ | $Z_7$ | $Z_8$ | $\hat{X}_1$ | $\hat{X}_2$ | $\hat{X}_3$ | $\hat{X}_4$ |
| $X_1$ | 12.631 | | | | | | | | | | | |
| $X_2$ | 0.422 | 12.591 | | | | | | | | | | |
| $X_3$ | 0.625 | 0.328 | 12.445 | | | | | | | | | |
| $X_4$ | 0.327 | 0.640 | 0.367 | 12.125 | | | | | | | | |
| $Z_5$ | 0.411 | 0.260 | 0.404 | 0.290 | 5.333 | | | | | | | |
| $Z_6$ | 0.324 | 0.279 | 0.405 | 0.305 | 0.222 | 5.463 | | | | | | |
| $Z_7$ | 0.293 | 0.361 | 0.241 | 0.411 | 0.186 | 0.271 | 5.603 | | | | | |
| $Z_8$ | 0.300 | 0.501 | 0.286 | 0.519 | 0.336 | 0.230 | 0.295 | 5.381 | | | | |
| $\hat{X}_1$ | 0.514 | 0.471 | 0.521 | 0.517 | 0.800 | 0.631 | 0.570 | 0.584 | 6.490 | | | |
| $\hat{X}_2$ | 0.428 | 0.565 | 0.416 | 0.604 | 0.460 | 0.494 | 0.639 | 0.886 | 0.833 | 7.115 | | |
| $\hat{X}_3$ | 0.502 | 0.441 | 0.533 | 0.482 | 0.758 | 0.760 | 0.452 | 0.536 | 0.977 | 0.781 | 6.636 | |
| $\hat{X}_4$ | 0.438 | 0.564 | 0.425 | 0.606 | 0.479 | 0.504 | 0.678 | 0.858 | 0.865 | 0.998 | 0.797 | 7.342 |

注：+表示对角线上输入的是标准差，对角线之外输入的是相关系数，所有变量的均值为0。

资料来源：原初变量的相关资料来自 Duncan 等人（1971：222），原初变量的标准差来自 Hanushek 和 Jackson（1977：280），工具变量由本书作者建构。

## 附录4：抗议事件中暴力行为模型方程5.13至方程5.15的参数估计中使用的所有变量的二元积矩相关矩阵

| | 原初变量+ | | | | | | | | | 工具变量+ | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | $X_1$ | $X_2$ | $Z_3$ | $Z_4$ | $X_5$ | $Z_6$ | $Z_7$ | $Z_8$ | $Z_9$ | $\hat{X}_1$ | $\hat{X}_2$ |
| $X_2$ | 0.61 | | | | | | | | | | |
| $Z_3$ | 0.28 | 0.33 | | | | | | | | | |
| $Z_4$ | 0.37 | 0.39 | 0.24 | | | | | | | | |
| $X_5$ | 0.62 | 0.51 | 0.44 | 0.56 | | | | | | | |
| $Z_6$ | 0.52 | 0.42 | 0.27 | 0.16 | 0.29 | | | | | | |
| $Z_7$ | 0.39 | 0.29 | 0.09 | 0.11 | 0.17 | 0.44 | | | | | |
| $Z_8$ | −0.39 | −0.30 | −0.11 | 0.01 | −0.26 | −0.29 | 0.09 | | | | |
| $Z_9$ | 0.41 | 0.38 | 0.18 | 0.46 | 0.40 | 0.24 | 0.24 | −0.03 | | | |
| $\hat{X}_1$ | 0.70 | 0.60 | 0.40 | 0.53 | 0.58 | 0.74 | 0.56 | −0.56 | 0.58 | | |
| $\hat{X}_2$ | 0.69 | 0.62 | 0.54 | 0.63 | 0.64 | 0.68 | 0.47 | −0.58 | 0.61 | 0.98 | |
| $\hat{X}_5$ | 0.58 | 0.57 | 0.63 | 0.80 | 0.70 | 0.42 | 0.24 | −0.37 | 0.57 | 0.83 | 0.92 |

注：+表示所有变量的均值为0，所有变量的标准差为1。

资料来源：原初变量的相关资料来自 Kritzer（1977：638），工具变量由本书作者建构。

# 附录 5：符号标记对照表

| 符号标记 | 意　　义 |
| --- | --- |
| cov(X，Y) | 变量 X 和 Y 的总体协方差 |
| var(X) | 变量 X 的总体方差 |
| $S_{XY}$ | 变量 X 和 Y 的样本协方差 |
| $S_X$ | 变量 X 的样本方差 |
| E(X) | 变量 X 的期望值(即总体均值) |
| $\hat{X}$ | X 的工具变量 |

# 注释

[1] 决定个人对选举中候选人偏好的因素的相关文献极其丰富。如读者有意深入了解该领域，可参阅 Goldberg(1966)、Jackson(1975)、Page 和 Jones(1979)。

[2] 关于"因果关系"意义的讨论，可参阅 Asher(1983:8—13)或者 Blalock(1969)的著作。

[3] 我将采用社会科学文献中最典型的称谓"因果图示"来指代变量之间因果关系的假设。在这些图示中，从变量 X 到变量 Y 的箭头(X→Y)表示 X 是 Y 的原因。

[4] 一个无偏估计值或参数值是指这样一个数值，其均值等于该参数的真实值，即如果 $E(\hat{\theta})=\theta$，那么 $\hat{\theta}$ 是 $\theta$ 的无偏估计。也可见注[27]。更详细的讨论可参考 Wonnacott 和 Wonnacott(1979:55—58)的著作。

[5] 我们在传统的统计学意义上使用"一致"这个词。概言之，如果样本数量趋向无穷，$\hat{\theta}$ 的分布趋向所有概率集中于 $\theta$ 的分布，那么 $\hat{\theta}$ 就是 $\theta$ 的一致性估计值。更详细的讨论参见 Christ(1966:263—264)。

[6] 有关一般最小二乘法回归分析的介绍，可参考 Lewis-Beck(1980)的著作。有关递归因果模型的讨论，可参考 Asher(1983:30—48)或 Duncan(1975:第 1—4 章)的著作。

[7] 这是社会科学文献中对前定变量的一个传统定义，一个更精确的定义是，前定变量与模型中所有误差项皆不相关。

[8] 当非递归模型中包含作为前定变量的滞后内生变量时，我们必须非常谨慎。详细解释请参见附录 1。

[9] 本书的讨论局限于线性叠加模型。对线性和叠加性的讨论参见 Lewis-Beck(1980)的专著。为了表述的简便，我忽略了结构方程中的常数项。但常数项也可内含于方程 1.2 中，只要把其中一个外生变量(如 $Z_1$)的值设为 1。

[10] 如果我们同时假定方差齐性，即某递归模型中的误差项具有相同的方差 [$VAR(\varepsilon_j)=\sigma_j^2$]，高斯-马尔科夫法则就可以确保系数的估计值是最优线性无偏估计值(BLUE)，即在所有的线性无偏估计值中，该估计值具有最小的方差(参见 Wonnacott & Wonnacott, 1979:21—28)。

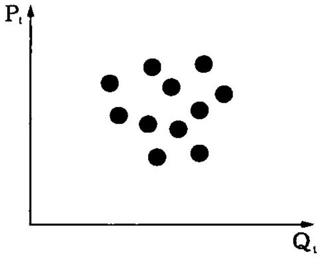
[11] 我们还假定模型中所有变量和误差项的均值为 0，这仅仅是为了通过调节变量的测量尺度的原点来简化方程的表征，以去除方程中的常数项。

[12] 为了简化小麦市场模型的分析，我们一直假定 $D_t$ 和 $P_t$ 之间、$S_t$ 和 $P_t$ 之间的关系是决定性的，也就是说，每个 $P_t$ 值都对应于唯一的 $D_t$ 和 $S_t$ 值。但我们也可以通过在方程 2.5 和方程 2.6 中加入误差项建立一个随机模型，这也更符合典型的社会科学研究：

$$Q_t = a_D + b_D P_t + \varepsilon_D$$

$$Q_t = a_S + b_S P_t + \varepsilon_S$$

然而如前所述，这个市场模型方程组依然是不可辨识的。如果这个模型是市场中供给和需求动力运作的真实写照，那么在一段时间内，对小麦出售量和价格的观察就会产生一组点数据，它们聚集在真实的供给和需求曲线交汇处周围，当两条曲线随机偏离它们的真实位置时，这些点就由每一时期误差项的大小决定：




我们仅仅观察到这些点是不足以精确估计需求和供给曲线的参数的。

[13] 参考附录 2 的说明，但更详尽的说明请参阅 Fisher(1966)的著作，他讨论了产生方程可辨识性的其他假设限定种类。

[14] 不可否认的是，"最小限度充分"这一概念过于模糊。我将在"简化式视角"一节中专门讨论方程是适度辨识还是过度辨识的必要条件。

[15] 既然 $U = \varepsilon_1$，按照模型的原初假设，U 与 $Z_3$ 就不相关。同样，既然 $Z_3$ 被假定为与 $\varepsilon_1$ 和 $\varepsilon_2$ 都不相关，而 V 等于一个常数项($\beta_{21}$)与 $\varepsilon_1$ 相乘再加上 $\varepsilon_2$，所以 $Z_3$ 与 V 必然不相关。

[16] 这些有关线性组合的证明，请参阅 Christ(1966:315—318)和 Goldberger(1964:312—313)的著作。

[17] 然而，如果并非所有变量的均值为 0，那么方程中就存在常数项，但这些常数项在可辨识性检验时可以被忽略。

[18] 熟悉线性代数的读者可以在 Wonnacott & Wonnacott(1979:第 18 章)和 Hanushek & Jackson (1977:254—264)的著作中找到更传统(但也是等价的)对秩条件的介绍。

[19] 如果结构方程包含常数项，那么在形成系统矩阵时，可以不考虑它们。

[20] 我知道从这一步开始，该运算法则中某些个别步骤背后的意图一点也不清晰。但大多数读者并不需要了解每一步运算背后的详细逻辑。重要的是要记住，整体来看，运算法则的每一步都是在模拟一个多方程模型中被选方程的线性组合过程。

[21] 熟悉线性代数的读者会发现，所谓的"简单式"就是"行简化式"，因为星号代表非零的数。

[22] 一个合理的平行假设是，受访者朋友的智力也是他们期望水平的原因。但这一假设与图 3.2 所示的模型不符，因为 $cov(\varepsilon_2, \varepsilon_4)$不再假定为 0 了。如果该假设为真，朋友的智力水平就成为导致误差项 $\varepsilon_2$ 和 $\varepsilon_4$ 相关的一个因素。

[23] 该模型由 Hanushek 和 Jackson(1977)改编自 Duncan 等人的研究。

[24] 对于这一修正模型，$m = k = 4$。对于方程 4.1 和方程 4.2，$k_e = 2$，$m_i = 2$，且运算得出 $R = 3$，所以对于这两个方程，$R = m - 1$ 且 $k_e > m_i - 1$。对于方程 4.3 和方程 4.4，R 也等于 3，且 $k_e = 2$，但 $m_i = 3$。

[25] 该模型的 $m = 3$，$k = 5$。对于 3 个结构方程而言，$m_i = 3$ 且 $R = 2$。对于方程 4.6 和方程 4.7，$k_e = 3$，但对于方程 4.5，$k_e = 2$。

[26] 需要注意的是，给模型添加一个外生变量永远都不会使先前可辨识的方程变得不可辨识，但它可能会使一个适度辨识的方程变得过度辨识。

[27] 有关期望的含义，变量协方差与变量乘积的期望值等价以及变量方差与变量平方的期望值等价的证明，可参见 Asher(1983:附录 A)的著作。

[28] 当某个估计值的抽样分布的方差比其他估计值小时，这个估计值就被认为更有效。更具体的讨论参见 Wonnacott 和 Wonnacott(1979:58—60)的著作。

[29] 更详细的完全信息估计方法的讨论，可参见 Christ(1966)的著作。

[30] 有关完美多元共线性所带来的后果的详尽讨论，请参阅 Lewis-Beck(1980:58)的著作。

[31] 有关处理高多元共线性其他方法的一般讨论，请参阅 Lewis-Beck(1980:58—63)的著作。

[32] 有关 2SLS 中的多元共线性问题更为详细的讨论以及这一影响的几个说明，请参见 Kritzer (1976)的著作。

[33] 附录 3 提供了期望数据所有变量的标准差以及所有变量之间的相关。

[34] 该模型的 $m = 3$，而所有方程的 $R = 2$。方程 5.13 的 $k_e = 5$，$m_i = 3$；方程 5.14 的 $k_e = 1$，$m_i = 2$；方程 5.15 的 $k_e = 5$，$m_i = 2$。

[35] 抗议事件暴力行为模型中的变量的大多数指标都是由"事件问卷"中的几个项目来测量的（参见 Kritzer，1977）。附录 4 提供了分析所用到的所有变量的相关矩阵。

[36] 本书的 2SLS 估计与 Kritzer(1977)所报告的估计值稍有不同，这可能是由复制 Kritzer 的分析所使用的相关矩阵中的化整误差造成的。

[37] Kritzer(1976:87)认为，系数 $\beta_{15}$ 的一个负值是"荒谬的"，而我相信负值与威慑理论是吻合的（如 Gamson，1975），该理论认为逮捕行为将消解抗议者进一步的暴力行为。

[38] 可以预计，一个多方程模型是否存在不可观测变量，很大程度上取决于有关误差项的假定。

# 参考文献


Achen, C. H. (1977). "Measuring representation: perils of the correlation coefficient." *American Journal of Political Science 4*:805—815.

Asher, H. B. (1983). *Causal Modeling*. Beverly Hills, CA: Sage.

Blalock, H. M., Jr. (1969). *Theory Construction*. Englewood Cliffs, NJ: Prentice-Hall.

——(1976) "Causal inference, closed populations, and measures of association." *American Political Science Review 61*:139—136.

Christ, C. F. (1966). *Econometric Models and Methods*. New York: John Wiley.

Deegan, J. (1975). "The process of political development: an illustrative use of a technique for regression in the presence of multicollinearity." *Sociological Method and Research 3*:384—415.

——(1966) "Path analysis: sociological examples." *American Journal of Sociology 72*:1—16.

——A. O. Haller, & A. Porter(1971). "Peer influences on aspirations: a reinterpretation." In H. M. Blalock, Jr. (ed.), *Causal Models in the Social Sciences*. Chicago: Aldine-Atherton.

Erikson, R. S. (1976). "The influence of newspaper endorsements in presidential elections." *American Journal of Political Science 20*:207—223.

Fisher, F. M. (1966). *The Identification Problem in Econometrics*. New York: McGraw-Hill.

Gamson, W. A. (1975). *The Strategy of Social Protest*. New York: Dorsey Press.

Goldberger, A. S. (1966). "Discerning a causal pattern among data on voting behavior." *American Political Science Review 60*:913—922.

Goldberger, A. S. (1964). *Econometric Theory*. New York: John Wiley.

—— & O. D. Duncan(eds.)(1973), *Structural Equation Models in the Social Sciences*. New York: Seminar Press.

Hanushek, E. A. & J. E. Jackson(1977). *Statistical Methods for Social Sciences*. New York: Academic Press.

Hibbs, D. E., Jr. (1973). *Mass Political Violence: A Cross-National Causal Analysis*. New York: John Wiley.

Hoerl, A. & R. Kennard(1970). "Ridge regression: biased estimation for nonorthogonal problems." *Technometrics 12*:55—67.

Jackson, J. E. (1975) "Issues, party choices, and presidential votes." *American Journal of Political Science 19*:161—185.

Klein, L. R. (1962). *An Introduction to Econometrics*. Englewood Cliffs, NJ: Prentice-Hall.

Kritzer, H. M. (1977). "Political protest and political violence: a nonrecursive causal model." *Social Forces 55*:630—640.

——(1976). "Problems in the use of two stage least squares." *Political Methodology 3*:71—93.

Lank, K. C. (1971). "Significant others, the self-reflexive act and the attitude formation process: a reinterpretation." *American Sociological Review 36*:1085—1098.

——(1969). "Principles of path analysis." In E. F. Borgatta and G. W. Bohrnstedt(eds.), *Sociological Methodology*. San Francisco: Jossey-Bass.

Lemieux, P. (1978). "A note on the detection of multicollinearity." *American Journal of Political Science 22*: 183—186.

Lewis-Beck, M. A. (1980). *Applied Regression: An Introduction*. Beverly Hills, CA: Sage.

Long, J. S. (1983). *Covariance Structural Models: An Introduction to LISREL*. Beverly Hills, CA: Sage.

Mason, R. & A. N. Halter(1968). "The application of a system of simultaneous equations to an innovation diffusion model." *Social Forces 47*: 182—195.

Namboodiri, N. K., L. E. Carter & H. M. Blalock, Jr. (1975). *Applied Multivariate Analysis and Experimental Designs*. New York: McGraw-Hill.

Ostrom. C. W. (1980). *Time Series Analysis: Regression Techniques*. Beverly Hills, CA: Sage.

Page, B. & C. Jones(1979). "Reciprocal effects of policy preferences, party loyalties and the vote." *American Political Science Review 73*: 1071 1089.

Sullivan, J. L. & S. Feldman(1979). *Multiple Indicators: An Introduction*. Beverly Hills, CA: Sage.

Theil, H. (1971). *Principles of Econometrics*. New York: John Wiley.

Tufte, E. R. (1974). *Data Analysis for Politics and Policy*. Englewood Cliffs, NJ: Prentice-Hall.

Waite, L. J. & R. M. Stolzenberg(1976). "Intended childbearing and labor force participation among young women: insights from nonrecursive models." *American Sociological Review 41*: 235—251.

Wonnacott, R. J. & T. H. Wonnacott(1979). *Econometrics*. New York: John Wiley.

# 用面板数据做因果分析

## 作 者 简 介

### 史蒂芬·E. 芬克尔(Steven E. Finkel)

美国维吉尼亚大学政府与国际事务系副教授。在《美国政治科学评论》、《美国政治学杂志》、《政治学杂志》、《公共舆论季刊》等学术期刊上发表了很多关于政治参与、公共态度以及定量研究方法的文章。

## 译 者 简 介

### 李　丁

北京大学社会学系本科和硕士毕业,现为北京大学社会学系博士研究生。

法，Finkel在书中都进行了认真的介绍。另外，与自相关有关的某些假定必须满足以使模型得到识别。除了这些问题外，测量误差问题对面板数据的影响尤为严重。正如Finkel所证明的，当这一因素被考虑时，(模型的)结论可能大不相同。为了演示如何处理测量误差问题，他从简单的单指标、二期模型一直讲到更复杂的模型，如三指标、三期模型。

很重要的一点是，本书让读者明白，只有满足了一些假定，才能进行有意义的参数估计。另外，为了方便研究实践，书中提供了大量的真实数据应用案例。在这一点上，有两个数据集用得恰到好处，一个是1987年至1989年前联邦德国抗议行为的跟踪调查，另一个是1980年美国大选跟踪调查。总之，在假定读者已掌握回归分析及LISREL用法和功能的基础上，Finkel博士为我们提供了清晰易读的、先进的面板分析技术。

迈克尔・S.刘易斯-贝克(Michael S. Lewis-Beck)

# 第 1 章 | 导论

面板数据在社会科学中通常被用来对个体变化及社会变迁理论进行检验，这种数据由来自相同个体或单位在几个时点上的信息构成。和静止的截面分析相比，面板分析最重要的特征在于，它将变化本身明确地包含在设计中，从而使个体（或其他层次的单位）在一系列变量上的变化得到直接测量。面板数据应该和另外两种形式的纵贯数据区别开(Menard,1991)："重复截面"或"趋势"数据，这种数据由在不同的时点上向多个不同单位收集的相同变量信息组成；另外一种是"时间序列"数据，这种数据所含的观察结果来自同一个个体在不同时点上的多个变量(Ostrom, 1978)。面板数据的不同之处在于，它们包含了来自多个被重复观察的单位在相同变量上的测量数据。[1]

本书对适用于面板数据分析的各种模型进行了概览，特别关注了跟踪调查优于截面研究设计的主要领域：变量间因果关系的分析。众所周知，变量 $X$ 与 $Y$ 之间若要存在因果关系，则必须满足如下条件(Menard, 1991:17)：(1)$X$ 和 $Y$ 必须共变，在非实验研究中表现为存在不等于 0 的双变量相关；(2)从时间上看，$X$ 必须在 $Y$ 之前；(3)这种相关不是"虚假"的，或者说，并不是由 $X$ 和 $Y$ 与第三个变量或其他一系列变量的相关造成的。成功的因果推论还必须基于对所关注变量的准确测量，当观察变量的随机或非随机测量误差未被考虑和处理时，因果作用的统计估计结果就可能是不正确的(Berry & Feldman, 1985; Carmines & Zeller, 1979)。截面数据尽管能提供第一个相关条件上的证据，但它们在提供时间先后关系、相关非虚假性的证据以及在设置对变量测量误差进行纠正的模型方面的作用有限得多。我们后面将清楚地看到，面板数据在这些方面具有绝对优势。

在截面分析中，对变量的单时点测量使时间先后顺序的确立非常困难，从而无法排除这种可能：$X$ 和 $Y$ 之间的共变关系是由于 $Y$ 导致 $X$，或是通过相互因果关系造成的。相反，面板分析中对相同单位的 $X$ 和 $Y$ 的历时测量，使研究者设置某种必定满足时间先后标准的模型——每个变量的早前取值会对其他变量后来的取值造成影响——成为可能。另外，一旦怀疑变量间存在相互因果关系，面板分析能在假定限制条件比在截面分析中少的情况下，估计出变量间带反馈效应的非递归模型。面板分析在控制可能使 $X$ 和 $Y$ 之间的关系部分或完全成为虚假关系的外部变量的影响方面也非常有用。在截面分析中，只有通过在统计模型中实际纳入外部变量，才能对虚假关系进行检验，但在面板分析中，即便某些因不可测量的因素造成的虚假模式也能被检验。最后，在面板数据中拟合测量误差模型比在截面数据情况下拟合测量误差模型所必要的限制性假定少，分析者可利用新增的历时观察对变量的因果作用和测量属性进行估计。在上述所有方面，跟踪调查设计（面板设计）允许对因果关系进行比在截面数据下更严格的检验，因此，总的来说，它比其他观察研究设计更接近于在实验法下才可能进行的、对于因果关系的控制性检验。

本书将讨论两种使用面板数据进行因果分析的互补性观点。书中将呈现面板数据强化因果推论过程的多种方式，接下来的章节将演示如何估计各种含有不同时滞设定、交互影响及测量不完美变量的模型。与此同时，本书强调面板数据并不是非实验研究设计下解决因果推论问题的万灵丹。本书将表明，所有方法和模型都依赖于它们各自的一套需要在给定条件下被论证和合理化的假定。如果这些假定不堪一击或可能导致不合理的经验性结论，研究者想要对分析得出的因果结论怀有信心，就必须对其他替代模型进行估计和比较。

自从 Markus 的著作《面板数据分析》(*Analyzing Panel Data*)出版以来，有关面板分析的文献已有巨大增长。本书试图突出不同学科及不同分析传统的学者在面板数据分析上作出的贡献。但有些主题无法在一本书中得到充分覆盖。例如，本书仅集中关注分析定距因变量的线性模型，而 Markus(1979：第 2 章)介绍了离散型面板数据的分析技术，Plewis(1985：第 6—7 章)对新近的研究做了概括。本书的讨论主要限于面板模型的估计和解释问题，暂不讨论与面

板案例流失和死亡相关的问题以及与跟踪调查受访者的自我选择有关的偏差问题。关于这些主题的出色讨论，读者可以参看 Kessler 和 Greenberg(1981:第 12 章)以及 Menard(1991:第 4 章)的作品。

如果读者想从后面的展示中得到最大收益，就要对多元回归和因果建模方法有所了解(Asher, 1983; Berry, 1984; Berry & Feldman, 1985)。同时建议读者对更一般的结构方程方法，如 LISREL 或 EQS 等有所了解，因为这些程序现在已经是面板分析的常规工具了。有关 LISREL 的介绍，可以参考 Long (1983a、1983b)；关于 LISREL 和其他方法更为具体的讨论，可以参看下列文献：Bollen (1989)、Bentler (1985)、Hayduk (1987)、Jöreskog 和 Sörborm (1976)、McArdle和 Aber(1990)以及 McDonald(1980)。在本书有些章节，我们会用LISREL框架来展示和讨论一些模型，以方便应用这些方法，但这些模型同样可以用 EQS 或其他结构方程软件进行估计。书后的附录对 LISREL 符号系统做了概要介绍。

# 第2章 | 用面板数据对变化进行建模

面板数据含有每个个体或单位的各个变量在 $t$ 及 $t-1$、$t-2$ 等时点上的测量结果，时点多少取决于观察的轮数/期数。因此，我们有可能通过使用各个变量早前及当下的取值信息来建构和估计因果模型。滞后（延迟[①]）因变量 $Y$ 或"滞后内生"变量的存在允许我们直接分析 $Y$ 的历时变化，如果我们能够表明变量 $X$ 与 $Y$ 的变化（$\Delta Y$）有关，那么由此提供的 $X$ 与 $Y$ 之间存在因果关系的证据要比从静态的截面设计中可能获得的证据直接得多。此外，面板数据同时含有 $X$ 的当下值和滞后值，使我们可以对 $X$ 和 $Y$ 之间的因果作用进行各种不同的设置。除了估计 $X_{t-1}$ 对 $\Delta Y$ 的影响外，研究者还可以创建一个变量，以代表 $X$ 在不同调查轮次间的变化，然后将 $Y$ 或 $\Delta Y$ 建模为 $X_{t-1}$、$X_t$、$\Delta X$ 或这些变量的某种组合的函数。

不过，尽管 $X$ 和 $Y$ 的当下值及早前值的同时存在，为研究者分析（事物的）变化提供了重要的新信息，但在面板模型中设置这些变量的作用的确切方法并非如此显而易见。在本章，我们将看到，在大多数情况下，较好的面板分析模型只是通过因变量滞后值 $Y_{t-1}$ 及一系列自变量 $X$ 来估计因变量当下值 $Y_t$ 的"稳定得分"或"条件性变化"方程的某种变体。接着，我们将告诉大家是否选择 $X_t$、$X_{t-1}$ 和/或 $\Delta X$ 作为自变量，主要取决于各轮观察之间的间隔长度以及对 $X$ 到 $Y$ 的滞后性因果作用的性质及时间性的理论假定。最后，在后面几章里，我们会更具体地讨论估计和解释面板模型时的几个潜在问题。

---

① Lagged，翻译为"滞后"、"延宕"、"延迟"。指早前发生的事情或变量对后来的事情或变量具有延续性的影响。具有这种影响的变量被称为"滞后变量"或"延宕变量"。——译者注

# 变化得分模型与滞后内生变量的角色

## “无条件的”变化得分模型

使用面板数据估计自变量 $X$ 对因变量 $Y$ 变化的影响的可能办法之一，从对截面分析中常用模型的拓展开始：

$$Y_t = \beta_0 + \beta_1 X_t + \varepsilon_t \qquad [2.1]$$

其中 $Y_t$ 和 $X_t$ 为某个体或案例在时间 $t$ 上的因变量及自变量取值，$\varepsilon_t$ 为误差项。要在这一方程中纳入其他被观察的自变量并不会遇到新麻烦，这里给出只含两个变量的简单模型是为了讲解的方便。面板数据中，同一个变量会在一个以上的时点被测量，这些新增测量可以为模型参数的估计提供重要的信息。假定自变量在两次测量之间发生了一定程度的变化，且 $X$ 与 $Y$ 之间的因果过程与在时间 $t-1$ 时的相同，也就是说 $\beta_{1t-1} = \beta_{1t}$，那么，从方程 2.1 中减去一个使用时点 $t-1$ 的 $X$、$Y$ 和 $\varepsilon$ 值表示的类似方程即可得到：

$$Y_t - Y_{t-1} = (\beta_{0t} - \beta_{0t-1}) + \beta_1 (X_t - X_{t-1}) + (\varepsilon_t - \varepsilon_{t-1})$$

或者 [2.2]

$$\Delta Y = \Delta\beta_0 + \beta_1 \Delta X + \Delta\varepsilon$$

该方程表示 $Y$ 的变化对 $X$ 的变化的简单回归，被称为面板分析的“无条件的变化得分法”或“一阶差分法”（Allison，1990；Liker，Augustyniak & Duncan，1985）。

这一方程在几方面优于对应的截面方程 2.1。首先，通过在分析中使用真实的变化得分，该方程直接对变量在个体或单位层面的变化的决定因素进行建模，不像在截面分析中，自变量的“变化”对因变量的“变化”的回归估计完全基于同一时点上不同单位间的变异。这一优势是所有面板模型都具备的。更确切地说，无条件的变化得分法控制了方程 2.1 中某些被遗漏的解释变量或被计量经济学称为“个体固定效应”的东西。如果真实的解释模型包括某些未知或

因为某种原因无法被纳入模型的恒定自变量($Z$)，如下列方程所示：

$$Y_t = \beta_0 + \beta_1 X_t + \beta_2 Z + \varepsilon_t \quad [2.3]$$

那么，对 $\beta_1$ 的截面估计将会有偏，偏差程度与 $X$ 和 $Z$ 的相关程度一致(因为 $Z$ 被整个归入了方程 2.1 的误差项)。

不过，如果假定这些被遗漏的 $Z$ 变量对 $Y$ 的影响保持不变，那么，经过差分过程，它们会从方程 2.2 中被完全消除：

$$\Delta Y = \Delta\beta_0 + \beta_1 \Delta X + \beta_2 \Delta Z + \Delta\varepsilon$$
$$\Delta Y = \Delta\beta_0 + \beta_1 \Delta X + \Delta\varepsilon \quad [2.4]$$

由于方程 2.4 中 $Z$ 的混淆影响已被移除，自变量 $\Delta X$ 不再和误差项 $\Delta\varepsilon$ 相关，因为方程 2.3 中对应的 $X_t$、$X_{t-1}$ 被假定与 $\varepsilon_t$ 及 $\varepsilon_{t-1}$ 不相关。因此，方程 2.4 能得到 $X$ 对 $Y$ 的因果作用的无偏估计，而方程 2.1 和方程 2.4 估计的 $\beta_1$ 之差就代表了截面模型方程 2.1 由于遗漏相关恒定变量 $Z$ 而导致的错误设定偏差的大小。正因为这种特性，“即便模型只被部分设定，面板数据分析仍能对模型中自变量的影响进行一致的估计”(Arminger, 1987:339)。这无疑是面板模型相对截面模型的一个明显的优势。

不过，尽管一阶差分模型或无条件的变化得分面板模型在估计此种错误设定的模型上很有用，但它含有一个极为严格的假定：滞后因变量(或“滞后的内生变量”)$Y_{t-1}$ 对 $Y_t$ 和 $\Delta Y$ 都没有影响。后面我们将看到，这一假定可能是不正确的，正因为如此，将无条件变化模型作为结构模型来分析变化通常是错误的。[2]

## 稳定得分或条件变化模型

将滞后因变量纳入方程 2.1 即得到所谓的稳定得分或条件变化面板模型(Plewis, 1985)，如下所示：

$$Y_t = \beta_0 + \beta_1 X_t + \beta_2 Y_{t-1} + \varepsilon_t \quad [2.5]$$

在这一模型中，$Y_t$ 由其早前值 $Y_{t-1}$、同期自变量 $X$ 及一个被假定方差恒定、不

存在自相关、不与 $X_t$ 或 $Y_{t-1}$ 相关的误差项来解释。当 $X$ 不恒定且在多个时点得到测量时，$X_{t-1}$ 甚至 $X$ 的其他滞后值的影响同样可被纳入模型，这会在后面再进行讨论。

这一模型同样可以用 $\Delta Y$——因变量的历时变化——来表达，只要从方程 2.5中减去 $Y_{t-1}$ 就可得到：

$$\Delta Y = \beta_0 + \beta_1 X_t + (\beta_2 - 1)Y_{t-1} + \varepsilon_t \qquad [2.6]$$

从这一设定可以看到，在稳定得分模型中，$X$ 对 $Y$ 的因果作用 $\beta_1$ 可被解释为控制因变量初始值后，$X$ 对 $\Delta Y$ 的因果作用，而方程 2.6 中 $Y_{t-1}$ 对 $\Delta Y$ 的影响就是方程 2.5 中 $Y_{t-1}$ 对 $Y_t$ 的稳定效应减去 1。

前面讨论的无条件变化模型也可以用 $Y_t$ 和 $Y_{t-1}$ 进行表达，通过将 $Y_{t-1}$ 移到方程 2.4 的右边并限定其因果作用等于 1 即可：

$$Y_t = \Delta\beta_0 + \beta_1 \Delta X + Y_{t-1} + \Delta\varepsilon \qquad [2.7]$$

有时，无条件变化模型被认为只是稳定得分公式的一种被限定形式(Hendrickson & Jones, 1987)。这是不正确的，因为 $Y_{t-1}$ 必然与方程 2.7 中的误差项 $(\varepsilon_t - \varepsilon_{t-1})$ 相关，但 $Y_{t-1}$ 被假定与方程 2.5 中的误差项 $\varepsilon_t$ 无关(Allison, 1990:103)。这意味着两个模型在设定上存在根本差别，对分析者来说，或估计方程 2.4 的无条件变化模型，或估计方程 2.5 或方程 2.6 的稳定得分模型。在非实验研究中有不少理由支持我们选择稳定得分模型。

## 稳定得分模型的实际合理性

首先，有大量的理由可以假定 $Y_{t-1}$ 是 $Y_t$ 或 $\Delta Y$ 的原因之一。在政治、社会或心理的态度分析中，早前倾向会对倾向的当前状态或其历时变化产生一定的因果作用。例如，对现任总统某个月的表现表示赞赏的人，至少会部分因为他们此前的态度而更有可能对总统下个月的表现表示支持。另一个例子是，一个人以往的收入不仅是当前收入的良好预测变量，并且确实对当前收入具有因果作用，因为有钱人采取的投资策略带来的收入增长比穷人所做的经济决定带来的增长大得多(Plewis, 1985:59)。在官僚主义决策制定模型中，通常假定一个

机构以往各年的预算或支出会对当年的开支额度产生一定的因果影响。一般而言，只要因变量的当前状态（或因变量的变化）直接由以往状态所决定，那么在这种情况下，要想正确设定模型，就必须纳入滞后因变量。相反，如果变量在每个时期都必须“重生”，那么将早前值纳入模型作为当前状态的预测因素也就没有什么实质基础了。[3]

### 向均值回归

即便没有明确的实际理由支持滞后 $Y$ 变量的纳入，稳定得分模型从统计上来讲也具有合理性。其原因在于，遗漏滞后 $Y$ 变量将无法顾及变化分析中最常见的现象之一：变量初始值与其后续变动之间极可能存在的负相关或者通常所说的“向均值回归”。由于某个时点，$Y$ 取值较大的个体或单位在后续时点的取值将偏小，而取值较小的个体，后续取值将偏大。对于这一倾向的忽略，将使无条件变化得分模型估计结果有偏，偏差程度与解释变量 $X$（或 $\Delta X$）和 $Y$ 的初始值的相关程度一致。

“回归效应”导致面板数据中 $Y_{t-1}$ 与 $\Delta Y$ 负相关的原因很多。原因之一是 $Y$ 的随机测量误差的存在。因为 $Y_{t-1}$ 取值较大的原因之一可能是测量误差较大，而这种误差在下一轮调查时将趋于变小。在极端情况下，$Y$“事实上”根本没有变化，全部的观察变化都是由测量误差造成的，而且可以看到，只要测量误差方差一直保持不变，$Y_{t-1}$ 与 $\Delta Y$ 之间的协方差就会等于 $Y$ 的误差方差的相反数（Dwyer，1983：339）。另外，“向均值回归”在测量完美的情况下也能存在。因为 $Y_{t-1}$ 取值极端，部分是因为较大的误差项 $\varepsilon_{i-1}$（代表所有被遗漏变量及纯随机因素的影响）造成的，而误差项在下一轮调查中趋向变小，所以 $Y$ 的变化很可能与 $Y_{t-1}$ 负相关。如果情况真是这样，那么忽略 $Y_{t-1}$ 会导致对所有同时与 $Y_t$ 及 $Y_{t-1}$ 正相关的自变量 $X$ 对 $\Delta Y$ 的影响的低估。

回归效应在面板数据中并非总是出现，但已经证明一旦下述条件得到满足，一个变量的初始值与其后续变化之间的负相关就会出现：(1)该变量各个时期的取值并非完美相关；(2)其方差相对固定（Bohrmstedt，1969；Kessler & Greenberg，1981；Nesselroade，Stigler & Baltes，1980）。在此类情况下，将

$Y_{t-1}$纳入回归模型是控制此种现象的方式之一，并可对分析作出如下解释：在$Y_{t-1}$的固定水平上，也就是说，在考虑了$Y$的初始值与后续变化之间的负影响后，自变量$X$是否会影响$Y$的变化？不过，在第4章中我们将看到，如果回归效应是因$Y_{t-1}$的测量误差造成的，那么模型还需要进一步修正。

## 负反馈作用

在面板模型中纳入滞后$Y$变量的另一个理由来自社会系统的稳定性概念。如果一个因果系统会在未来某个时期达到一个固定平衡点使每个案例的$Y$取值保持不变，除非系统本身被某种外部干扰因素所改变，那么它就被认为是"稳定的"(Arminger, 1987; Dwyer, 1983)。虽然经验研究所分析的绝大多数系统尚未达到平衡，但可以证明系统的稳定性要求$Y_{t-1}$对$\Delta Y$存在"负反馈"效应(Coleman, 1968)。如果$Y_{t-1}$对$\Delta Y$的影响是正的，那么$Y$将无限扩展：如果$Y_{t-1}$为负值，那么随着时间的变化，$Y$将越来越负；如果$Y_{t-1}$为正数，那么$Y$会越来越正。无论何种情况，$Y$的方差都会不断"爆炸"，这在绝大多数社会—心理系统中，都被认为是不可能发生的(尽管在技术上并非不可能)。因此，$Y_{t-1}$对$\Delta Y$的负向影响会引导系统趋向平衡：当$Y$超出平衡水平时，它就会降低；当$Y$低于它的平衡水平时，它就会提高。

$Y_{t-1}$对$\Delta Y$的此种负反馈也被解释为某些通过模型而被遗漏的、将$Y_{t-1}$与$Y_t$连接起来的因果路径的代理变量。Coleman(1968)声称，$Y_{t-1}$对$Y_t$的正向影响(因而$Y_{t-1}$对$\Delta Y$的负效应)可以被看做：

> 经验系统中尚不为正式(因果)系统所知道的所有反馈链条的代理者。当正式系统更加完善时，这一系数应该趋近于0。因此，该系数的大小可以用来评估对经验系统作出表征的完善程度。(1968:441)

从这个角度来看，将因变量的早前水平纳入模型，至少可以部分控制影响$Y_t$变化的被遗漏的变量。不过，作为这种角色，滞后的$Y$具有一种不同的"认识论地位"，而非一个实际意义上对$Y_t$构成直接因果作用的变量，

其估计效应的解释也因此需要相应调整。(Arminger, 1987; Liker et al., 1985)

## 部分调整

滞后 $Y$ 值被纳入某些面板模型,是为了估计一些其他理论上相关的参数。例子之一是最先在经济学中流行起来的部分调整模型。在这种模型中,需要用自变量进行解释的是某种未知的、"可欲的"、"最优的"或"目标"值($Y_t^*$)而非因变量的实际值($Y_t$),因此,其实际方程为:

$$Y_t^* = \beta_0 + \beta_1 X_t + \varepsilon_t \qquad [2.8]$$

$Y_t^*$ 常常被看做前面已经描述过的 $Y$ 的均衡水平,但渴望值 $Y^*$ 还可代表其他目标,如一个组织的目标。用理性行动的术语来说,就是能够给个体带来最大效用的 $Y$ 值(Tuma & Hannan, 1984:339)。

根据这一模型,个体或组织一直力求使 $Y$ 与 $Y^*$ 之间的差异最小化,但 $Y$ 的实际变化只等于 $Y_t^*$ 与 $Y_{t-1}$ 差异的一部分 $\alpha$。也就是说,由于惯性、无知或者阻止变迁的结构性因素等影响,每个时期只能"部分地调整"$Y$ 的渴望值与实际值之间的差距。这一想法可以表述如下:

$$Y_t - Y_{t-1} = \alpha(Y_t^* - Y_{t-1}) \qquad [2.9]$$

其中,系数 $\alpha$ 表示调整系数,或者说是从 $t-1$ 到 $t$ 期间,渴望值与实际值之间的差距被缩小的程度。将方程 2.8 中的 $Y_t^*$ 值代入方程 2.9,即得到估计方程:

$$Y_t - Y_{t-1} = \alpha\beta_0 - \alpha Y_{t-1} + \alpha\beta_1 X_t + \alpha\varepsilon_t \qquad [2.10]$$

这一方程与条件变化模型方程 2.6 有着同样的一般形式——以 $Y_{t-1}$ 及 $X_t$ 来解释 $\Delta Y$。从方程 2.10 中可以看到,$Y_{t-1}$ 对 $\Delta Y$ 的回归作用等于调整参数 $\alpha$ 的相反数,估计的作用越接近于一1,在给定时间内,调整的 $Y$ 越接近其"渴望"值或平衡值。从估计方程 2.10 中得到的 $X_t$ 的回归作用有两种解释方式:从初始形式来看,它代表调查轮次间 $X_t$ 对 $Y$ 或 $\Delta Y$ 的短期作用;将这一取值除以 $\alpha$ 即得到方程 2.8 中的 $\beta_1$,后者代表 $X_t$ 对平衡值或渴望值 $Y_t^*$ 的长期作用。因此,部

分调整模型提供了一种不同但具有互补性的理由，以支持在面板模型中纳入 $Y_{t-1}$ 作为 $\Delta Y$ 的解释变量。[4]

## 稳定得分模型的估计

我们将用来自 1987 年至 1989 年前联邦德国跟踪调查的数据示范说明稳定得分模型。当初进行这一调查是为了模拟和解释政治抗议活动中的个体参与（关于这一研究及其抽样过程的更多细节，请参阅 Finkel, Muller & Opp, 1989）。这里，我们感兴趣的两个变量在两轮调查中都得到测量，一个是个体合法抗议潜能指标（$PROTEST_1$ 及 $PROSTEST_2$）的对数得分，这一指标是用个体过去 8 种非暴力行为的参与情况对某些未来行为意向组合进行加权测量得到的，这些非暴力行为包括为请愿书收集签名和参加合法游行示威等；另一个是群体成员资格指标（$GROUPS_1$ 及 $GROUPS_2$），这一变量代表受访者从属的、他们自己认为会鼓励抗议行为的群体组织的数量。时间 $t-1$ 时的抗议潜能可能通过一个部分调整过程与时间 $t$ 时的抗议潜能连接起来，在此，$Y_t^*$ 代表使个体效用最大化的抗议水平，而群体成员资格变量（$GROUPS_2$）则代表社会压力或群体动员过程的一个部分，正是通过这一过程，个体从参与中得到效用。由于这一原因，同样也因为上面已经描述过的统计原因，我们设定了稳定得分模型以代表 $X$ 和 $Y$ 之间的历时关系。在这一章中，我们假定 GROUPS 和 PROTEST 之间的因果关系是单向的。

如果稳定得分模型中的误差项被假定与 $X_t$ 和 $Y_{t-1}$ 都不相关，那么模型系数通过一般最小二乘（OLS）回归就可得到一致的估计。在这种假定下，OLS 通常产生的估计不仅一致（即在 $N$ 趋近无穷大的过程中，估计结果将趋近于总体真值）且无偏（即来自随机样本的无穷多个估计，既不会高估也不会低估真实参数），但是，当滞后因变量被纳入递归模型时，OLS 估计将含有一个随着 $N$ 变大而趋近于 0 的偏差（Johnston, 1972）。因此，存在滞后因变量的 OLS 估计仍具有一致性，但从技术上来讲并非无偏。本例来自 OLS 估计的结果如下：

$$\text{PROTEST}_2 = 0.22 + 0.10\ \text{GROUPS}_2 + 0.40\ \text{PROTEST}_1 \qquad [2.11]$$
$$(0.02)\ (0.02) \qquad\qquad (0.04)$$
$$\underline{0.30} \qquad\qquad \underline{0.40}$$

括号中的数字为标准误，标准化系数为带下划线的数字，模型的确定系数 $R^2$ 为 0.34。这一结果表明，控制因变量的原初水平后，每多一个个体在 1989 年参加群体组织，将使 1989 年抗议潜能的对数增加 0.10，且该作用在常规水平上统计显著。这一模型也可以这样解释：在控制原初抗议水平的情况下，1989 年的群体资格每增加一个，将导致抗议潜能在 1987 年到 1989 年间变化 0.10。1987 年的抗议潜能对 1989 年的抗议潜能的稳定作用是 0.40，这一效应同样在统计上显著。与之等价的解释是，1987 年的抗议潜能对两轮调查间抗议水平的变化的作用是(0.40－1)或－0.60。从部分调整的观点来看，$\alpha$ 等于 0.60，这意味着从一个时期到下一个时期，抗议水平被其最优(平衡)水平调整了 60%。从长期因果作用来看，群体成员资格指标每增加一个单位，对抗议潜能的最优或平衡水平的影响为 0.17(0.10 除以 0.60，即根据方程 2.10，将估计的 $\beta_1$ 除以 $\alpha$)。

## 其他滞后设定

刚才估计的模型规定自变量 GROUPS 对同一时点的 PROTEST 具有影响，也就是说，$X_t$ 被假定导致 $Y_t$。事实上，我们不能说 $X$“即刻”造成了 $Y$，而只能说，相对于两轮测量之间的时间间隔，$X$ 影响 $Y$ 的因果时滞很短。[5]但这种“同步”或“同时”效应设定只是多种可供试探的设定之一，通过跟踪调查数据对这些设定进行试探，可以确定 $X$ 对 $Y$ 的因果作用。如果 $X$ 对 $Y$ 构成影响的时滞间隔充分长，但仍短于两轮调查之间的间隔，那么滞后作用模型可能更合适，在这种模型中，$Y_t$ 是 $Y_{t-1}$ 和 $X_{t-1}$ 的函数。如果 $X$ 一直保持高度稳定，不同的设定选择并不会带来太大差异，因为这与从 $X_{t-1}$ 或 $X_t$ 对 $Y_t$ 进行因果作用估计得到的结果相似。但如果 $X$ 在不同调查轮次之间发生显著变化，那么由因果时滞设置不当所可能造成的偏差将相当大。

为变量之间的作用设定合适的时滞结构是面板分析(及其他纵贯分析)中最难的问题之一,之所以如此,原因大致有二:第一,研究者通常不确定在理论上,$X$ 需要花多长时间才能对 $Y$ 造成影响。如果一个人参加了某种具有抗议取向的群体或次级组织,到底需要经过 1 天、1 个星期、3 个月,还是 6 个月,其抗议倾向才能被强化呢?第二,即使知道理论上的因果时滞,也无法保证这一间隔与观察轮次之间的间隔一致,因为管理的方便性、资金的可获得性之类的因素和实际关心的因素一样,决定着追踪调查的测量间隔。由于这种不确定性,面板分析中,时滞结构的设定首先应该在理论指导下进行,其次才以特定研究情境下的经验证据为依据。

**离散时间面板模型**

$X$ 和 $Y$ 之间的因果滞后关系的重要方面之一是,变化究竟是"跳跃式"地离散发生,还是持续不断地进行。例如,议会选举结果,每隔一个固定的两年就变化一次,而对总统的评价在为期 4 年的任期内几乎是连续变化的。连续时间过程的面板模型由 Coleman(1968)发展而来,最近又得到了 Tuma 和 Hannan(1984)的发展,我们将在下一节对这些方法进行更详细的描述。如果对象的变化被假定是离散地发生的,那么时滞结构的设置首先应该基于一种事先预期——$X$ 究竟需要花多长时间才能对 $Y$ 产生影响?同时要结合考虑跟踪调查各轮测量之间的间隔。如果缺乏明确的理论指导对模型时滞长度进行恰当的设定,分析者可以尝试以经验的方式决定因果时滞。

例如,在后面的章节中,我们将就非法伙伴群体涉入情况对青少年自身越轨行为的影响进行分析。在这个例子中,$X$ 和 $Y$ 之间的因果时滞应该相对较短,希望一个青少年 2 至 3 年前的伙伴群体影响其现在的行为可能是不合理的。因此,如果每隔 3 年才收集一次面板数据,"滞后作用"模型就不是合适的设定,而"同步"效应模型能更好地描述违法伙伴对青少年违法犯罪的因果作用。不过,如果隔 1 或 2 个月就收集一次数据,那么滞后作用设定会更合适,因为这一间隔与伙伴群体中的社会压力、计划等因素促成违法行为实际所需的时间更对应。

在理论上，我们可能要求一个模型同时具有滞后作用和同步作用。请试想，压抑或痛苦的人生经历对个体当前的心理健康造成的影响。若跟踪调查周期只有1年或2年，那么有理由认为当前痛苦的生命事件会对同期的心理健康造成影响。同时，2年以前的生命事件($X_{t-1}$)也会对当前的心理健康具有挥之不去的直接影响，或通过一些不可测量的变量，如个体的身体状态、工作业绩之类的因素，对当前健康形成间接影响(Kessler & Greenberg, 1981:78—79)。

在单向因果模型中，同时纳入 $X_t$ 和 $X_{t-1}$ 不会带来估计上的严重问题(除变量非常稳定时可能出现的高度共线性问题外)，因此，如下形式的模型可被估计，从而为寻找恰当的滞后因果关系提供一些线索：

$$Y_t = \beta_0 + \beta_1 X_t + \beta_2 Y_{t-1} + \beta_3 X_{t-1} + \varepsilon_t \qquad [2.12]$$

这一方程在直观上更好地代表了自变量的变化与因变量的变化之间的关系。通过使用等式 $X_t = X_{t-1} + \Delta X$，这一方程的系数可用 $X$ 的变化重新表述。方程2.12可以表达为：

$$Y_t = \beta_0 + \beta_2 Y_{t-1} + (\beta_1 + \beta_3) X_{t-1} + \beta_1 \Delta X + \varepsilon_t \qquad [2.13]$$

请注意，方程2.13中 $\Delta X$ 对 $Y_t$ 的效应 $\beta_1$ 与方程2.12中 $X_t$ 对 $Y_t$ 的效应相同。换言之，$X$ 的变化对 $Y_t$ 有影响即意味着，在控制了 $X$ 和 $Y$ 的初始值后，$X_t$ 对 $Y_t$ 有影响(Kessler & Greenberg, 1981:10)。

方程2.13的参数可以通过这些代数运算得到，也可通过将 $X_1$ 和 $\Delta X$ 作为解释变量纳入回归模型而直接得到(这还能得到对应的标准误)。类似的运算同样可以用来将方程2.12变为 $\Delta X$ 和 $X_2$ 的表达式。[6]这些模型在代数上是等价的，因此，结果的实质性解释取决于模型的理论假定。如在政治稳定模型中，不稳定程度很可能被假定与一个国家当前的经济水平($X_2$)负相关，但正如快速经济增长是“失稳力量”这一假设所示，它也可能受到一段时间内经济水平变化($\Delta X$)的正向影响。而在政治竞选对投票者的影响模型中，可能假定个体投票行为同时取决于某些初始特征(例如，对现任管理者的表现的认同程度 $X_1$)和选举期间的突发事件造成的认同程度的变化($\Delta X$)(Finkel, 1993)。因此，将 $X$ 的当前值与滞后值作为 $Y$ 的变化的预测变量纳入其中的方程可以有多种解释方

式，具体取决于模型的实际关怀。

我们对群体资格—抗议行为例子中的稳定得分模型进行了再次估计，将1987年的群体资格变量 $GROUPS_1$ 纳入，作为1989年的抗议倾向的解释变量。从理论上讲，如果群体发动的抗议行为在事前相当长的时间内就已计划展开，那么群体成员资格对抗议行为的影响就会存在一定程度的滞后。但是为期2年的观察间隔相对于滞后间隔可能又太长了，因而同步作用模型可能更好。OLS估计结果支持这一观点。用 $X_{t-1}$ 和 $Y_{t-1}$ 解释 $Y_t$ 的纯滞后作用模型表明，GROUPS对PROTEST的作用比方程2.13代表的同步模型小得多：

$$\underset{}{PROTEST_2} = \underset{(0.02)}{0.22} + \underset{\underline{0.11}}{\underset{(0.02)}{0.04}}\ GROUPS_1 + \underset{\underline{0.47}}{\underset{(0.05)}{0.47}}\ PROTEST_1 \qquad [2.14]$$

括号中是标准误，标准化系数为带有下划线的数字，$R^2$ 为0.28，比同步模型中的小。估计一个同时含有 $GROUPS_1$ 和 $GROUPS_2$ 作为自变量的稳定得分模型，结果如下：

$$PROTEST_2 =$$

$$\underset{(0.02)}{0.22} + \underset{\underline{0.31}}{\underset{(0.02)}{0.11}}\ GROUPS_2 + \underset{\underline{-0.02}}{\underset{(-0.02)}{-0.01}}\ GROUPS_1 + \underset{\underline{0.40}}{\underset{(0.04)}{0.40}}\ PROTEST_1 \qquad [2.15]$$

括号中为标准误，标准化系数为带有下划线的数字，$R^2$ 为0.34。这一结果表明，在控制了 $Y$ 变量及群体资格两者的滞后影响后，当前的群体成员资格与抗议倾向显著相关。控制当前群体成员资格变量后，滞后群体成员资格变量在统计上不显著，并且对模型解释力的贡献不大。根据方程2.13，$GROUPS_2$ 的影响同样可以解释为，在控制滞后的群体成员资格变量及滞后的抗议倾向变量后，群体成员资格数的变化对 $PROTEST_2$ 的影响。

## 连续时间面板模型

上一节讨论了在既定的（调查）时间间隔下，估计 $X$ 的作用落实到 $Y$ 上的恰

当时滞结构的方法。如果理论时滞长度与观察轮次间的时长对应,滞后作用模型就较合适;如果时滞长度比各轮调查间的间隔短得多,那么同步作用模型会更好。不过,在有些面板模型中,$X$ 对 $Y$ 的影响几乎可以看做是连续地发生的,而不是在各轮观察之间离散地跳跃发生。例如,个体的党派身份对他们对总统候选人的态度的影响在整个竞选过程中持续存在,因为随着时间的变化,个体会根据其内在的政党忠诚度不断调整对主要政党候选人的评价。企业组织多少会不断调整其雇用基数以应对环境中的经济力量。此类模型直观上及理论上都更适合分析社会科学家通常研究的很多因变量,如态度及其他社会—心理概念、组织变迁、人口流动等等。在这种情况下,特定时点所进行的测量代表的只是用以观察因果过程的纯粹人为间隔内的变化,分析者的任务是要发现 Coleman 所说的"变化的基本参数"。下面即可看到,估计连续时间模型所用的方法与离散时间模型类似,但对参数的解释显著不同。

我们通过模拟 $Y$ 对时间 $t$ 的导数来表达基本的连续时间模型:

$$\frac{dY_t}{dt} = c_0 + c_1 X_t + c_2 Y_t \qquad [2.16]$$

$Y$ 对 $t$ 的导数这一概念来自代表 $Y$ 的瞬间变化速率——也就是说,$t$ 变化无穷小时 $Y$ 的变化量——的初级计算公式。根据这一方程,变化量的大小同时取决于给定时点上的 $X$ 值和 $Y$ 值。还有很多其他模型可以用来预测 $Y$ 的变化速率(例如,将速率预测为常数或时间的函数等),但方程 2.16 最具理论吸引力,它将 $Y$ 的瞬间变化速率作为 $Y$ 及其他被假定对 $Y_t$ 具有因果作用的变量的函数。

要估计这一模型,解或"求导"这一微分方程就能将 $Y$ 的当前值表达为 $t-1$ 到 $t$ 期间,方程 2.16 中各变量的累积效应。暂且假定外生变量 $X$ 为恒定变量,求导方程 2.16,得到如下结果:

$$Y_t = \frac{c_0}{c_2}(e^{c_2 \Delta t} - 1) + \frac{c_1}{c_2}(e^{c_2 \Delta t} - 1)X + e^{c_2 \Delta t} Y_{t-1} \qquad [2.17]$$

其中 $e$ 为自然对数,$\Delta t$ 为各轮调查间的时间间隔。如果在方程 2.17 中加上一个随机误差项以概括 $Y_t$ 的其他原因,那么这一方程可被看做方程 2.5 稳定得分

模型的另一种形式，以 $X$ 和 $Y_{t-1}$ 作为自变量对 $Y_t$ 进行预测：

$$Y_t = \beta_0 + \beta_1 X + \beta_2 Y_{t-1} + \varepsilon_t \qquad [2.18]$$

与离散时间模型中的回归系数不同，连续时间估计模型中的 $\beta_k$ 代表的是方程 2.17 中系数 $c$ 和测量间隔的复杂非线性函数。不过，只要满足关于残差项的标准假定，且没有设定错误或测量误差，$\beta_k$ 就可以通过 OLS 进行估计，而上面的系数 $c$——对应"变化的基础参数"——可以这样反求出来：

$$c_0 = \beta_0 \frac{\ln \beta_2}{\Delta t(\beta_2 - 1)}$$

$$c_1 = \beta_1 \frac{\ln \beta_2}{\Delta t(\beta_2 - 1)} \qquad [2.19]$$

$$c_2 = \frac{\ln \beta_2}{\Delta t}$$

## 一个例子

我们将用 1980 年美国全国选举跟踪调查数据来说明上述方法，这一数据包含了对随机抽中的 763 名符合选举年龄的成年人的调查结果，调查分别在 1980 年的 1 月、6 月、9 月和 11 月进行。有关总统选举期间的态度变化的研究显示，对于候选人的态度受到个体政党忠诚倾向的直接影响，因而高度认同共和党或者民主党的个体，对自己政党推举的候选人的评价在竞选结尾阶段比开始时更为积极，但对反对党候选人的评价，随着时间推移会越来越差(Markus, 1982)。在该研究中，受访者被要求在一个 100 分的"感觉温度计/民意评分"量表上，对两个主要的政党候选人——共和党的罗纳德・里根(Ronald Reagan)及民主党的吉米・卡特(Jimmy Carter)——打分，在此基础上构建出一个变量，代表对两个候选人的打分差异，使正数得分表示对里根的评价更高，而负数得分表示对卡特的评价更高。在下面的分析中，我们将 6 月这一变量称为 $THERM_1$，而 9 月为 $THERM_2$。我们用 6 月的一个 7 分政党认同量表作为自变量，取值从"坚定的民主党员"一直到"坚定的共和党员"。下面的系数、标准误及标准化系数

是从方程 2.18 的 OLS 估计中得到的：

$$\text{THERM}_2 = -2.05 + \underset{\substack{(0.54)\\ \underline{0.20}}}{3.74}\ \text{PID}_1 + \underset{\substack{(0.03)\\ \underline{0.65}}}{0.66}\ \text{THERM}_1 \qquad [2.20]$$

$$\quad (0.99)$$

两个自变量的系数都统计显著，模型总的确定系数 $R^2$ 为 0.59。估计结果为 1980 年竞选过程中"（候选人）民意评分"的党派极化假设提供了支持。政党认同度朝支持共和党方向每增加 1 个单位，与个体 6 月到 9 月间对候选人的相对态度评分向支持里根（反对卡特）的方向变化 3.74 个单位显著相关。在控制政党认同后，6 月的评价水平对 9 月的评价水平的效应为 0.66，也就是说，这一模型表明，6 月的评价得分对 6 月至 9 月间评价得分变化的作用是 −0.34(0.66 − 1)。

从连续时间的角度来看，方程 2.20 中的系数可用于求出潜在微分方程中的系数 $c$，这一系数描述了政党认同对候选人评价变化的瞬时速率的影响。根据方程 2.19 的相等关系，如果假定 $\Delta t$ 取值为 3 与两轮调查间为期 3 个月的间隔对应，那么候选人评分($Y_{t-1}$)对 $Y$ 相对于时间 $t$ 的导数的作用为−0.14，而政党认同($X_{t-1}$)的作用为 1.53，截距 $c_0$ 的估计结果为 0.88。

这些系数给出了有关这一因果系统的重要信息。第一，系统是稳定的，也就是说，由于存在系数 $c_2$(−0.14)所代表的"负反馈"作用，它最终会达到平衡。第二，特定案例的均衡 $Y$ 值，可以通过令导数式等于 0 来求解，即 $Y = (-c_0 - c_1X_1)/c_2$。因此，如果个体在选举开始时的政党认同得分为 3(对应"坚定的共和党人")，那么模型预测的均衡 $Y$ 值为民意差异量表得分 39.1；一个坚定的民主党人(政党认同得分为−3)的均衡值则为−26.5。第三，系数 $c_2$ 的大小表示 $Y$ 向其均衡值移动的速度。如果 $c_2$ 很小（比如值很大的负数），那么在来自 $X$ 的外生效应影响下，$Y$ 移向均衡点的速度将非常快。当 $c_2$ 变大时，$Y$ 趋向稳定状态的速度将不断变小。在这里，$c_2$ 取值−0.14 表明 $Y$ 趋向均衡的过程较慢。如果潜在的实际模型被看做一个部分调整过程，$Y$ 不断调整以趋向某一目的或期望的效应目标，那么可以证明调整参数 $\alpha$ 等于方程 2.16 中系数 $c_2$ 的相反数

(Tuma & Hannan，1984:337—338)。在此，$\alpha$ 估计值 0.14 表明 $Y$ 趋向渴望目标的速度非常慢，这和均衡论解释是一样的。

连续时间模型可以被扩展以处理外生变量在跟踪调查期间发生变化的情况。在这种情况下，要想反求出方程 2.16 中的系数 $c$，就必须对 $X$ 的历时变化模式作出一些假定。在 $X$ 的变化为线性模式的简单假定下，估计模型中增加另外一个因素($\Delta X$)后所得到的积分方程的解要在方程 2.17 中加上如下部分：

$$\frac{c_1}{c_2}\left[\frac{e^{c_2\Delta t}-1}{c_2\Delta_t}-1\right]\Delta X \qquad [2.21]$$

在这一方程中，$\Delta X$ 前面的数量即用 $Y_{t-1}$、$X_{t-1}$ 及 $\Delta X$ 预测 $Y_t$ 的回归模型的系数 $\beta_3$。

如果 $Y$ 到 $X$ 之间不存在反馈效应，并且误差项表现良好，这一模型同样可以用 OLS 方法进行估计。对于政党认同的例子，这种估计的结果为：

$$\begin{aligned} \text{THERM}_2 = &-1.61 + 4.33\ \text{PID}_1 + 0.64\ \text{THERM}_1 + 3.27\ \Delta\text{PID} \qquad [2.22]\\ &(-0.97)\ (0.23) \qquad (0.03) \qquad (0.96)\\ &\qquad\quad \underline{0.63} \qquad\quad \underline{0.23} \qquad\quad \underline{0.08} \end{aligned}$$

括号中的数字为标准误，带下划线的数字为标准化系数，$R^2$ 为 0.59。要将 OLS 系数转化为微分方程中的参数 $c$，就必须首先根据上面的方程 2.19 从 $\beta_2$ 估计出 $c_2$，然后利用其他两个回归系数产生 PID 效应 $c_1$ 的两个估计：一个来自类似方程 2.19 的系数 $\beta_1$，另一个来自系数 $\beta_3$：

$$c_1 = \frac{\beta_3 c_2^2 \Delta t}{\beta_2 - 1 - \ln \beta_2} \qquad [2.23]$$

在这个例子中，从系数 $\beta_1$ 估计的 $c_1$ 取值为 1.79，而从 $\beta_3$ 估计的 $c_1$ 取值为 2.57。不管是哪一个，都表明早前在外生变量固定不变的假定下得到的系数 1.53 低估了政党认同对候选人评分变化瞬时速率的真正影响。Tuma 和 Hannan(1984)讨论了将这两个取值组合起来并从数据中得到关于 $c_1$ 的多种最佳估计方法，从简单的代数平均值一直到混合的最大似然估计法。[7] 因此，连续

时间模型中 $X_{t-1}$ 及 $\Delta X$ 对应的两个系数被用来提供 $X$ 对 $Y$ 的“连续同步”影响的单一估计。

## 面板模型估计中的问题

上述讨论看起来会让人以为面板数据背景下的因果推论只是一个简单的问题：设定连续或离散的稳定得分模型，通过 OLS 回归对因果作用进行估计，然后将其解释为控制 $Y_{t-1}$ 后，$X$ 对 $Y_t$ 或 $\Delta Y$ 的影响。但即使是在最基本的稳定得分方程中，也存在几个妨碍因果推论成功进行的障碍，这些潜在问题经常使一般最小二乘法不适合作为面板分析的估计程序。最重要的问题来自如下可能的模型设定错误：

第一，互为因果。如果 $X_t$ 和 $Y_t$ 在一个交互因果系统中相互影响，那么 OLS 将产生有偏且不一致的参数估计，从而必须寻找其他需使用更多变量或更多观察轮次的估计量。

第二，测量误差。如果 $Y_{t-1}$ 未被准确测量，那么其对 $Y_t$ 的实际影响将无法得到确切估计，也就不适合作为 $Y$ 的前期水平与后续变化间常见的负向关系的恰当控制。确实，如上面讨论过的，$Y_{t-1}$ 的测量误差是 $Y_{t-1}$ 与 $\Delta Y$ 之间负相关的可能根源之一，但试图通过直接在模型中纳入 $Y_{t-1}$ 以纠正问题，又会因为自变量带有误差时的有关统计估计问题而引入新的偏差。这种偏差通常会低估 $Y_{t-1}$ 对 $Y_t$ 的影响，高估其他解释变量的影响。

第三，被遗漏变量或自相关的扰动项。最后，被遗漏变量可能给面板模型带来几种类型的偏差。除了由被遗漏变量与观察自变量之间的相关而导致的常规设定偏差外，面板模型中被遗漏变量还能导致外生变量不同时期的误差项之间的自相关。这又会使 $Y_{t-1}$ 和 $\varepsilon_t$ 之间产生非零相关，最终使稳定得分模型中 $Y_{t-1}$ 对 $Y_t$ 效应的 OLS 估计不一致。如果其他自变量与 $Y_{t-1}$ 相关，自相关扰动项也会使它们对 $Y_t$ 影响的估计有偏。

稳定得分模型中的这些设定与估计问题将成为本书剩余部分讨论的内容。尽管这些问题给成功的因果推论带来了巨大的障碍，但我们将看到，在

这些困难面前，跟踪调查数据往往能提供足够的信息以成功估计出这些参数。另外，这些问题中很多是在所有经验研究中都存在的，而且在后面我们将看到，和截面分析相比，跟踪调查数据为研究者控制这些问题提供了大得多的力量。

# 第 3 章 | 交互因果模型

上一章的模型都假定 $X$ 和 $Y$ 之间为单向关系，也就是说，只有 $X$ 影响 $Y$，不存在相反的影响。在部分情况下，这一假定完全适用。例如，在分析种族等先天特征对个体收入变化的影响模型中，或者模拟成年前期的经历对后来的政治及社会取向的研究中，变量之间的时间顺序（进而潜在的因果次序）是清楚的。在另一些情况下，理论原因将避免对交互因果关系的检验，例如，在试图模拟一组被长期观察的国家中经济指标对政府声望的影响的研究中就是这样。在这种模型中，$X_t$、$X_{t-1}$ 和（或）$\Delta X$ 可以被当做相应方程中的外生变量来对待，而参数估计可以通过 OLS 回归获得。再者，如果没有测量误差或自相关误差项的假定无法满足，可以通过后面的章节即将讨论的其他方法来获得。

不过在很多分析中，单向因果假定是站不住脚的，而且实际上，进行跟踪数据分析的主要原因之一就是为了确定所关注的变量之间的因果顺序。例如，在前面的章节中，我们假设群体成员资格会影响一个人的抗议行为倾向，而个体长期的政党认同会决定其在一次竞选过程对总统候选人的感觉，但是有关的参与及组织动员理论认为，参与抗议活动可能会导致个体参与更多具有抗议倾向的组织，而政治参与理论认为，对政治候选人的态度也会改变个体长期的政党认同。在这些情况下，不同理论将使我们怀疑 $X$ 和 $Y$ 之间存在交互因果关系。

相对于截面分析，面板数据在检验变量间可能存在的交互因果作用上具有绝对优势。由于截面数据是在一个时点上收集的，交互效应模型中只能设定一个变量对另一个变量的同步或同时因果作用；而且对交互因果作用的估计需要通过将外部变量整合进“工具变量”分析或二阶段最小二乘分析来进行（Berry，1984）。不过，下面我们将看到，这些方法成功与否取决于模型是否满足几个有

关这些外部变量与 $X$、$Y$ 及相应方程的扰动项的关系的严格假定。如在第 2 章中看到的，面板设计的历时特性允许研究者估计带滞后因果作用的模型，在这种模型中，$X$ 的早前值会影响 $Y$ 的未来取值(或 $Y$ 的变化)或者相反($Y$ 的早前值会影响 $X$ 的未来值)。另外，带有同步或同时因果作用的模型可以在不对外部工具变量的影响作出假定的情况下得到识别和估计，这些假定可能是靠不住的，但在截面研究中却是必要的。

本章将展现在交互作用模型中使用面板数据协助进行因果推论的主要内容。我们将强调面板设计是估计交互因果作用的有力手段，尽管它们无法提供自动"证明因果关系"的方法。交互因果关系的估计总是发生在特定的假设模型中，这些模型所包含的假定在特定情境下(是否成立)必须进行论证。具体来讲，二期数据交互效应模型所涉及的一系列可能的严格假定与在截面研究中一样，只有通过纳入外部变量才能被放宽。但三期或多期面板模型只需给因果参数较少的限制，就能被估计出来。

## 交叉滞后作用模型

### 二期数据模型

对可能的交互效应进行估计的最基本的模型是将上一章讨论过的滞后作用稳定得分模型(方程 2.14)扩展成两个因变量的形式，每个因变量在时间 2 上的取值都由其早前值及另一个相关变量在时间 1 上的取值来解释。该模型用路径图的形式呈现在图 3.1 中。$X_1$ 和 $X_2$ 分别代表一个变量(如群体成员资格数)在两轮调查中的取值，而 $Y_1$ 和 $Y_2$ 分别代表另一个变量(如抗议倾向)在两个时点上的测量结果。$X_2$ 和 $Y_2$ 被假设由它们各自在第一轮调查中的取值和另一个变量的滞后值以及误差项 $U$ 来决定。第一轮变量之间的相关由图 3.1 中的 $\rho_1$ 代表，第二轮方程中结构性扰动项之间的相关由 $\rho_{U_1U_2}$ 表示。两个结构方程可以写成如下形式：

$$Y_2 = \beta_1 X_1 + \beta_2 Y_1 + U_1$$
$$X_2 = \beta_3 Y_1 + \beta_4 X_1 + U_2 \quad [3.1]$$

上述模型中的所有变量都被表达成离均差的形式以消除截距项。

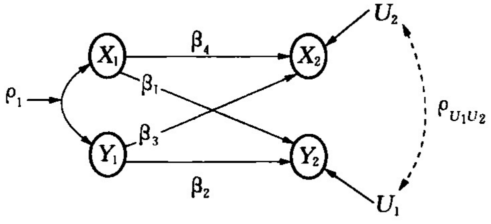



**图 3.1　带交叉滞后作用的二期数据模型**

交叉滞后模型在面板分析中用途广泛。当因变量的变化按离散时间过程被拟合时，只要因果时滞约等于各轮测量间的间隔，交叉滞后模型就是适用的。当因变量的变化连续发生时，可以证明交叉滞后模型代表一组微分方程的积分解，在该组方程中，$X$ 和 $Y$ 的瞬时变化速率相互决定，如下所示：

$$\frac{dY_t}{dt} = c_0 + c_1 X_t + c_2 Y_t$$
$$\frac{dX_t}{dt} = c_3 + c_4 Y_t + c_5 X_t \quad [3.2]$$

与上一章讨论过的单方程连续时间模型一样，方程 3.1 中的 $\beta_k$ 是方程 3.2 中的"变化基础参数"$c$ 以及各轮调查间的时间的非线性函数。从 $\beta_k$ 计算系数 $c$ 时，所涉及的复杂数学运算已经超出了本书的范围，感兴趣的读者可以从 Arminger (1986)、Coleman(1968:448—452)以及 Tuma 和 Hannan(1984:第 11—12 章)的著作中找到更详细的内容。就我们的目的而言，只需知道当因果系统是从一个变量到另一个变量的连续交互反馈系统时，交叉滞后模型"在因果作用的方向上一般是不会犯错的"(Dwyer，1983:352)就足够了。

另外可能注意到，这一交叉滞后模型对应于时间序列分析中所谓的因果关系的"葛兰杰检验"，后者断定在控制第二个变量早前取值的情况下，只要第一个变量在时间 $t-1$、$t-2$ 等等的测量值中，有任意一个对第二个变量在时间 $t$

的取值存在显著影响，那么第一个变量就是第二个变量的"葛兰杰原因"（Gujarati，1988：542—543）。在二期数据情况下，它被简化为方程 3.1 所示的模型，而在多轮跟踪数据下，所得为通过 $Y$ 和 $X$ 在时间 $t-1$、$t-2$ 等等的取值来预测 $Y$ 和 $X$ 在时间 $t$ 上的取值的交叉滞后模型。

## 交叉滞后模型的估计

起初，有人认为通过比较两个变量间的交叉滞后相关，即 $\rho X_1Y_2$ 和 $\rho Y_1X_2$，就可以确定变量间的因果方向。人们认为，如果以 $X_1$ 和 $X_2$ 为代表的变量对以 $Y_1$ 和 $Y_2$ 为代表的变量的因果作用比反方向的作用大，那么 $\rho X_1Y_2$ 将大于 $\rho Y_1X_2$。不过，后来无数学者的证明发现，这一推理过程是靠不住的，因为图 3.1 中变量间的交叉滞后相关是由好几个因素造成的：

$$\begin{aligned}\rho X_1Y_2 &= \beta_1 + \rho_1\beta_2 \\ \rho Y_1X_2 &= \beta_3 + \rho_1\beta_4\end{aligned} \qquad [3.3]$$

可以看到，交叉滞后相关不仅包含变量间的因果作用，还包括每个变量的历时稳定性以及第一轮时两个变量间的交互相关。因此，如果 $Y$ 的稳定性强于 $X$ 的稳定性，那么即使 $\beta_3$ 大于 $\beta_1$，$X_1$ 和 $Y_2$ 之间的交叉滞后相关也会大于 $X_2$ 和 $Y_1$ 之间的交叉滞后相关（Markus，1979：48—49）。由于因果作用的比较只有通过直接复原 $\beta_1$ 和 $\beta_3$ 才能进行，因此，回归及结构方程法已经不再将交叉滞后相关分析作为确定因果方向及强度的方法了。不过，在第 5 章中我们将看到，交叉滞后相关的比较对跟踪数据分析确实有一些用处，它可以检验某些变量间带有虚假相关的限制性模型。

在误差项 $U$ 的均值为 0、方差恒定，且与滞后外生变量 $X_1$ 和 $Y_1$ 无关的假定下，交叉滞后模型的参数可以通过一般最小二乘回归得到一致的估计。如前面已经提过的，这一模型中最让人感兴趣的是第一轮变量 $X_1$ 和 $Y_1$ 分别与 $Y_2$ 和 $X_2$ 之间的交叉滞后作用，因为它们被假定代表一个变量对另一个变量的因果作用。用上一章讨论过的变化得分语言来说，在控制 $Y_1$ 的情况下，$X_1$ 对 $Y_2$ 的影响代表 $X_1$ 对 $Y$ 的历时变化的影响；控制 $X_1$ 后的 $Y_1$ 对 $X_2$ 的影响可进行

同样的解释。通过 $X_2$ 和 $Y_2$ 方程的 $R^2$ 就能看出滞后变量对因变量的方差的解释程度，而误差项 $U_1$ 和 $U_2$ 之间的相关则代表未被模型中的交叉滞后（或连续时间）因果过程及稳定效应所解释的 $X_2$ 和 $Y_2$ 之间的共变。

将第 2 章中用群体资格数解释德国人合法政治抗议倾向的分析拓展一下，就可对这一方法进行示例说明。在前面，我们假定抗议倾向与群体成员资格之间是单向因果关系，但对图 3.1 所示的交叉滞后模型的估计结果显示，两个变量对对方都有显著的影响。这一模型的 OLS 估计结果（包括括号中的非标准化回归系数、标准误以及带下划线的标准化系数）为：

$$\begin{aligned}\text{PROTEST}_2 = 0.22 + &0.04\ \text{GROUPS}_1 + 0.47\ \text{PROTEST}_1 \qquad [3.4]\\ &(0.02) \qquad\qquad (0.05)\\ &\underline{0.11} \qquad\qquad \underline{0.47}\end{aligned}$$

其调整的 $R^2$ 为 0.28。另一个为：

$$\begin{aligned}\text{GROUPS}_2 = 0.01 + &0.58\ \text{PROTEST}_1 + 0.46\ \text{GROUPS}_1\\ &(0.14) \qquad\qquad (0.05)\\ &\underline{0.20} \qquad\qquad \underline{0.42}\end{aligned}$$

其 $R^2$ 为 0.29。两个方程中的回归系数都统计显著。这一结果表明，根据这一模型，第一轮调查时，个体拥有的鼓励抗议群体的成员资格数对两轮调查间抗议倾向的变化有影响；而第一轮调查时的抗议倾向对两轮调查间个体所属的鼓励抗议的群体数的变化也有显著的影响。事实上，抗议倾向对群体资格数的影响的大小是我们在第 2 章关注过的反向因果作用的将近两倍。模型中两个外生变量的结构扰动因素之间的相关为 0.22，与 1989 年 PROTEST 及 GROUPS 间观察到的相关系数 0.46 相比，这意味着直接观察到的两个变量间的相关的一半已被模型中设定的稳定效应及交叉滞后作用所解释。其他因素可能包括同步因果作用和与被遗漏外部变量的共同相关，可解释 $X_2$ 和 $Y_2$ 之间其余的共变。

### 三期或多期交叉滞后模型

当有三期或多期数据时，可使用之前讨论过的方法分别估计各个时期 $(t > 1)$ 的交叉滞后模型。不过，更为常见的办法是使用 LISREL 或相关结构

方程软件在整体模型中对这些系数进行同时估计，因为这些技术可以扩展以处理带有测量误差及自相关扰动项的复杂模型，并提供模型相对于其他设定情况的相对拟合优度，还允许对系数进行限定以估计我们关心的某些理论效应。

例如，现在我们有三期数据可用来估计 1980 年竞选期间，个体的政党认同($X$)与对总统候选人的评价($Y$)之间可能存在的交互因果关系。根据我们前面的讨论，如果假定这些变量之间的交互影响在竞选期间是持续地进行的，那么方程 3.2 刻画的微分方程所代表的因果系统里的参数可以通过交叉滞后模型的估计复原出来。我们将三期数据交叉滞后模型呈现在图 3.2 中，为了方便解释，我们使用 LISREL 符号系统(见附录)。

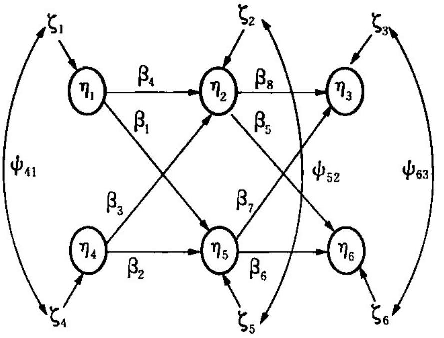



**图 3.2　带交叉滞后作用的三期数据模型**

第一轮(1980 年 2 月)、第二轮(6 月)、第三轮(9 月)调查中的政党认同分别由图中的外生变量 $\eta_1$、$\eta_2$ 及 $\eta_3$ 表示，而候选人(支持)民意得分由 $\eta_4$、$\eta_5$ 及 $\eta_6$ 代表。$\eta_1$、$\eta_4$ 实际上是事前已经确定的变量，不过在这里，我们将它们作为完全由它们的误差项决定的内生变量，这样，变量间所有的因果作用都处于 LISREL 内将所有内生变量连接起来的系数矩阵 **B** 中。这一设定不会对实际结论产生任何损害，但正如接下来将讨论的一样，只有这样，才能在各轮次间的 $\beta_k$ 上加上相等的限制。在 LISREL 设定中，这一模型只有两个相关的矩阵：**B** 矩阵和内生变量的结构扰动项 $\zeta_k$ 之间的方差协方差矩阵 $\mathbf{\Psi}$。$\zeta_1$ 和 $\zeta_4$ 之间的协方差 $\psi_{41}$ 代表变量 $\eta_1$ 和 $\eta_4$ 在第一轮调查时的协方差，$\zeta_2$ 和 $\zeta_5$ 以及 $\zeta_3$ 和 $\zeta_6$ 之间的协方差 $\psi$ 代表通过模型中设定的连续时间交互作用过程，在第二轮及第三轮调查中无法

解释的政党认同与候选人评分之间的共变。

如前面所讨论的，连续时间模型中的 $\beta_k$ 是各变量对 $X$ 和 $Y$ 的瞬时变化速率的影响以及各轮测量间的时间间隔的非线性函数。因此，如果各轮调查之间的时间间隔相同，那么各轮调查间的 $\beta_k$ 也应该一样，这样，从第一轮调查到第二轮调查间各变量之间的交叉滞后作用应该等于它们从第二轮调查到第三轮调查间的对应取值，就像各变量的稳定性估计一样。因而，给参数加上如下一致性或相等性限制：$\beta_1 = \beta_5$、$\beta_2 = \beta_6$、$\beta_3 = \beta_7$、$\beta_4 = \beta_8$，那么这一模型即可被识别。各变量影响效应不变的假定可通过放宽这些限定，然后比较前后两个模型的拟合优度得到检验。如果各轮调查之间的间隔不等，那么给不同轮次间 $\beta_k$ 的限制将是非线性的，而不是相等（Arminger，1987：343）。在 LISREL 框架中给不同轮次间的参数施以非线性限制尽管并非不可能，但却很难，读者可以从 Hayduk(1987)及 Rindskopf(1984)的著作中找到这些方法的应用案例。

为了示范，在当前这个例子里，我们假定结构效应是稳定不变的，尽管第一和第二轮调查之间的间隔比第二和第三轮之间长了大概一个月。另外，我们还假定 $\psi_{52} = \psi_{63}$，也就是说，第二轮调查中政党认同与候选人评分的扰动项之间的协方差与第三轮调查中的相等。其他所有误差项的协方差(除 $\psi_{41}$外)都被假定等于 0，这在本模型中，相当于假定各误差项不存在自相关。这些假定同样可以被放宽，并通过数据进行检验，更多细节将在第 5 章进行展示。在对结构性参数及扰动项施加了这些限定后，模型已经过度识别并带有 9 个自由度。该模型的LISREL最大似然估计结果在表 3.1 中显示。结果表明，政党认同与候选人评分之间存在显著的交互因果作用。政党认同对候选人评分的标准化影响接近反向作用大小的两倍。政党认同及候选人评分的稳定性都很高，模型得到的第二轮、第三轮数据中，政党认同及候选人评分的扰动项之间的标准化协方差取值很小，这意味着模型能够很好地解释这两个变量间的同步共变。

不过，可以看到，正如 9 个自由度下的 $\chi^2$ 值 201.9 所示，模型对数据的拟合非常差。将不同轮次的结构参数相等以及政党认同与候选人评分的扰动项之间的协方差在第二轮和第三轮中相等的假定放开后，得到的结构效应估计几乎

相同，对应的 $\chi^2$ 为 4 个自由度下的取值 190.2。由于第一个模型是“嵌套”在第二个中的，两个 $\chi^2$ 之差同样服从 $\chi^2$ 分布。在本例中，$\chi^2$ 之差为 10.9，对应的自由度为 5，这意味着不带相等性限制的模型在数据拟合度上并没有统计显著的改进。同时，由于整体拟合水平很差，模型还需要进一步的调整。在下一章中，我们将考虑放宽无测量误差及结构扰动项之间不存在自相关的假定。

**表 3.1　三期数据交叉滞后作用模型：政党认同与对候选人的评价（1980 年）**

| | 非标准化系数 | 标准化系数 |
|---|---|---|
| **稳定效应** | | |
| 候选人评价 | | |
| $\beta_2$ | 0.67[a] | 0.62 |
| $\beta_6$ | 0.67[a] | 0.64 |
| 政党认同 | | |
| $\beta_4$ | 0.83[b] | 0.82 |
| $\beta_8$ | 0.83[b] | 0.83 |
| **交叉滞后作用** | | |
| 政党认同对候选人评价 | | |
| $\beta_1$ | 3.98[c] | 0.21 |
| $\beta_5$ | 3.98[c] | 0.21 |
| 候选人评价对政党认同 | | |
| $\beta_3$ | 0.01[d] | 0.08 |
| $\beta_7$ | 0.01[d] | 0.09 |
| **误差协方差** | | |
| $\psi_{41}$ | 29.58 | 0.44 |
| $\psi_{52}$ | 4.45[e] | 0.06 |
| $\psi_{63}$ | 4.45[e] | 0.06 |
| $R^2(\eta_2)$候选人评价，第 2 轮 | 0.54 | |
| $R^2(\eta_3)$候选人评价，第 3 轮 | 0.58 | |
| $R^2(\eta_5)$政党认同，第 2 轮 | 0.73 | |
| $R^2(\eta_6)$政党认同，第 3 轮 | 0.77 | |
| $\chi^2$(df) | 201.1(9) | |

注：表中第一列为非标准化的最大似然估计系数，第二列为标准化系数。所有系数都统计检验显著。（原书中两种系数交替放在同一列，这里为了节省空间分成了两列。——译者注）

a、b、c、d、e：系数被限定相等。

资料来源：美国全国选举研究，主面板文件，1980 年。

## 同步作用模型

交叉滞后模型的估计方法相对简单，这无疑有利于使其成为流行的、处理因果关系的工具。但回忆上一章可知，这些模型中有关同步或同时效应为 0 的假定（也就是说，变量间的因果作用仅仅来自 $X_1$ 和 $Y_1$ 的滞后变量的影响）在很多研究情境下可能并不成立。当面板模型各轮测量的间隔相对于实际的因果时滞较长时，仅仅估计交叉滞后作用将无法完全刻画每个变量对其他变量的全部因果影响。另外，在有些模型中，交叉滞后作用的假定事实上是不成立的，如在 Dwyer 有关大学新室友与其他室友情绪的交互影响的例子中就是这样（Dwyer，1983:393）。变量间的同步或同时影响的出现，使其他估计交互因果作用的方法成为必需。

在有些情况下，一个适当的因果模型在特定时点上，各变量间只具有同步交互效应。在另一些情况下，变量间可能同时具有同步作用和交叉滞后作用，其中滞后值将解释变量部分而非全部的因果影响。同时包括滞后及同步作用的模型在识别和估计上会出现一些特殊问题，这个问题将会在介绍完相对简单的同步作用模型之后讨论。后者在两轮数据下的情形呈现在图 3.3 的示意图中，其对应的方程组为：

$$
\begin{aligned}
Y_2 &= \beta_1 Y_1 + \beta_3 X_2 + U_1 \\
X_2 &= \beta_2 X_1 + \beta_4 Y_2 + U_2
\end{aligned}
\qquad [3.5]
$$

$X_2$ 和 $Y_2$ 之间设定的交互作用意味着模型成为非递归模型，并给 OLS 估计带来了问题（Berry，1984）。因为 $X_2$ 与 $U_1$ 相关（后者通过 $Y_2$ 形成间接影响），$Y_2$ 与 $U_2$ 之间的关系相似，使与 OLS 中有关自变量与因变量的扰动项无关的假定被违反（如果总体中 $\beta_3$ 和 $\beta_4$ 不等于 0），因而 OLS 估计将对模型中的因果参数产生有偏估计。另外，很多非递归模型中的参数是无法识别的，因而我们必须确保观察到的方差协方差信息足够多，从而估计出感兴趣的系数。

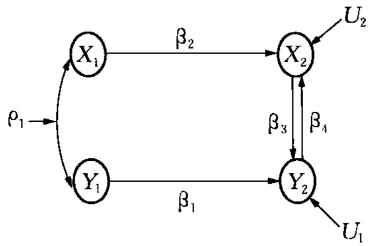



**图 3.3 带有同步作用的二期数据模型**

要识别和估计 $Y_2$ 方程，模型必须含有一个以上满足如下条件的"工具变量"：(1)与 $X_2$ 相关但与 $U_1$ 无关；(2)对 $Y_2$ 没有直接的因果作用。要识别和估计 $X_2$ 方程，至少需要一个与 $Y_2$ 相关但与 $U_2$ 无关，且对 $X_2$ 无直接因果关系的工具变量。在当前模型的假定下，滞后变量 $X_1$ 和 $Y_1$ 就满足这两个条件。因为它们被假定为是先定变量，所以从定义上讲，它们与两个方程的误差项无关；另外，由于模型假定两个交叉滞后作用都为 0，所以第二个条件同样得到满足。因此，两个方程都满足必需条件而恰好识别：两个各带有一个内生变量作为自变量的方程中，都有一个外生变量（技术上讲，先行确定的）被排除在外。能够利用滞后因变量来估计交互效应，是面板分析被公认的主要优势之一。不过需要注意的是，这种估计方法只有在满足严格的假定基础上才有效，包括不存在交叉滞后作用、连续的各轮调查的方程的误差项之间不存在自相关等。这些假定中只要有一个不满足，$X_1$ 和 $Y_1$ 就不能分别用来识别 $X_2$ 和 $Y_2$ 方程（Berry，1984）。

## 同步作用模型的估计与评价

由于同步作用模型中每个参数都恰好被识别，因此它们的取值可以通过对模型的 4 个"标准"方程，即两个内生变量 $X_2$ 及 $Y_2$ 分别与两个外生变量 $Y_1$ 和 $X_1$ 之间的预测协方差进行代数运算来估计。这些标准方程是这样得到的：将两个内生变量方程分别乘以每个外生变量，对所有案例全部加总，然后取期望值：

$$\begin{aligned}\mathrm{Cov}(X_1X_2) &= \beta_2\mathrm{Var}(X_1)+\beta_4\mathrm{Cov}(X_1Y_2)\\ \mathrm{Cov}(Y_1X_2) &= \beta_2\mathrm{Cov}(X_1Y_1)+\beta_4\mathrm{Cov}(Y_1Y_2)\\ \mathrm{Cov}(X_1Y_2) &= \beta_1\mathrm{Cov}(X_1Y_1)+\beta_3\mathrm{Cov}(X_1X_2)\\ \mathrm{Cov}(Y_1Y_2) &= \beta_1\mathrm{Var}(Y_1)+\beta_3\mathrm{Cov}(Y_1X_2)\end{aligned} \qquad [3.6]$$

$\beta_2$ 和 $\beta_4$ 由前两个标准方程计算得到，$\beta_1$ 和 $\beta_3$ 由后两个计算得到。

同样，这些参数也可以通过统计程序，如两阶段最小二乘(TSLS)从样本数据中估计出来，也能通过 LISREL 的最大似然估计或有关程序估计出来。不过，在此我们主要关注二阶段最小二乘法，以说明工具变量模型的运作逻辑。对于 $Y_2$ 方程，TSLS 法首先将 $X_2$ 对其工具变量 $X_1$ 及模型中所有其他先决变量进行回归，由此得到：

$$\hat{X}_2 = \beta_0+\beta_1Y_1+\beta_2X_1 \qquad [3.7]$$

请注意，所有先决(外生)变量，哪怕是 $Y_1$ 这样随后即导致 $Y_2$ 但并未明确设定为 $X_2$ 的原因的变量，都被用于第一阶段方程，目的在于得到 $X_2$ 最好的可能估计，因为第二阶段方程的系数的标准误直接取决于第一阶段估计的精度(Berry, 1984)。这是使用滞后变量作为工具变量的另一个优势，因为 $Y_1$ 和 $X_1$ 与它们在时间 2 上的取值的相关，很可能比和其他潜在外生变量的相关强。现在 $X_2$ 的估计值已与误差项 $U_1$ 无关，因为它是先定变量的线性组合，在定义上就与模型中所有的误差项无关。程序第二阶段将替代值 $\hat{X}_2$ 代入方程 3.5，所有的系数及对应的标准误都可以通过 OLS 得到一致的估计。两阶段过程被用来产生 $X_2$ 方程 3.5 中各参数的一致估计：将 $Y_2$ 对工具变量 $Y_1$ 及其他先定变量 $X_1$ 进行回归，产生一个与 $U_2$ 无关的 $\hat{Y}_2$，然后在第二阶段将 $X_2$ 对 $\hat{Y}_2$、$X_1$ 进行回归。上面讨论的假定，即 $X_1$ 对 $Y_2$ 无直接影响以及 $Y_1$ 对 $X_2$ 无直接影响，在这里非常重要，否则每个方程都会无法识别：$\hat{X}_2$ 和 $\hat{Y}_2$ 将成为它们对应的方程 3.5 中所有其他自变量的线性组合，从而无法得到因果作用的唯一估计。

尽管也可能把这些分析分成独立的两步进行，但目前的统计软件通常带有的两阶段最小二乘程序都是同时对两步中的系数进行估计。使用这些程序的

优点不仅在于计算方便，更重要的是它们还能提供正确的标准误、标准化回归系数以及第二阶段方程的 $R^2$ 等，这些程序中统计量的计算是基于 $X_2$ 和 $Y_2$ 的原始值而非预测值(Berry，1984：68—69)。用来自前联邦德国跟踪调查研究中的抗议倾向与群体成员资格的例子根据上述步骤计算，得到如下估计结果：

$$\text{PROTEST}_2 = 0.22 + \underset{\underline{0.26}}{\underset{(0.04)}{0.09}}\ \text{GROUPS}_2 + \underset{\underline{0.42}}{\underset{(0.06)}{0.42}}\ \text{PROTEST}_1 \qquad [3.8]$$

$R^2$ 为 0.30，另一个为：

$$\text{GROUPS}_2 = -0.27 + \underset{\underline{0.42}}{\underset{(0.29)}{1.24}}\ \text{PROTEST}_2 + \underset{\underline{0.37}}{\underset{(0.05)}{0.41}}\ \text{GROUPS}_1$$

$R^2$ 同样为 0.30。和交叉滞后模型中一样，两个方程的所有系数都在统计上显著。

乍一看，同步作用模型似乎比前面估计的交叉滞后模型更好。两个因变量被解释的方差比例都要大一些，抗议倾向与群体成员资格之间的交互因果影响更强。不过，同步作用模型一个很吸引人的特征在于，它还存在 1 个自由度，这可以用来检验模型的一个重要假定。尽管通过上文的介绍我们看到，在同步作用模型中，每个参数恰好被识别，但整个模型具有 1 个自由度，因为总共有 10 个观察的方差、协方差，但只有 9 个参数需要估计(4 个 β、$U$ 的两个方差、$X_1$ 和 $Y_1$ 的方差及协方差 $\rho_1$)。该自由度可以用来检验模型中有关内生变量的误差项之间的协方差为 0，即 $\rho_{U_1U_2}=0$ 的设定。这一残差协方差等于 0 的设定意味着，模型中的同步作用及变量的稳定性可以完全解释变量 $X_2$ 和 $Y_2$ 之间观察到的共变。在当前的例子中，$U_1$ 和 $U_2$ 之间估计的协方差为 −0.05，标准化取值为 −0.45，这一取值过大，实际上不太可能。[8]

尽管跟踪数据同步作用模型的扰动项之间可能存在负的协方差，但这样的结果解释起来比较麻烦。它意味着存在一个外部因素对这两个(正向)交互相关的内生变量中的一个具有正向影响，对另一个却是负向影响。这种因素实在

难以想象，更有可能的解释是模型设定不正确。Gillespie 和 Fox(1980)表明，很多同步作用模型的误差项之间出现负的协方差估计值可能与某些类型的模型设定失误有关，其中内生或外生变量存在测量误差、交叉滞后作用的排除、同时对外生及内生变量造成影响的变量的遗漏等，在面板模型中都意味着变量各时点的误差项之间存在自相关。除了内生变量的测量误差外，这些错误设定都会导致同步交互效应的计算往上发生偏误。如果不是因为残差项之间的负向共变所代表的“纠正”，模型将高估内生变量之间的共变。[9]因此，如果残差协方差为较大的负值，那么在同步模型中发现的正向交互效应可能只是模型设定错误的产物而已。在这种情况下，可以像下面讨论的那样，将交叉滞后作用纳入模型，而且/或者可以参照后面章节将讨论的方法对模型进行估计，这些程序可以对由测量误差和遗漏变量造成的可能偏差进行检验。

## 交叉滞后及同步作用模型

如果没有令人信服的理论原因来估计一个只带交叉滞后作用或同步作用的模型，或者同步作用模型所得到的结果不可能或无法解释，那么，同时带有两类效应的模型将是合理的替代设定。不过，在图 3.3 中加上从 $Y_1$ 到 $X_2$ 以及从 $X_1$ 到 $Y_2$ 的交叉滞后作用，将使模型有 11 个自由参数(需要估计)，但只有 10 个方差及协方差观察值。模型整体以及各个单独参数都将无法识别。这一识别问题在交叉滞后模型中是通过将同步作用限定为 0 来解决的，将交叉滞后作用限定为 0 才能使同步作用模型的参数被识别。假定两类效应都存在时，只有通过其他更多信息——或以外部变量的形式，或以更多调查期次的形式——才能使模型得到识别。

### 二期数据模型

当只有二期数据时，识别这一模型的唯一方法是纳入外部外生变量并将它们的某些影响设定为 0。要想识别 $Y_2$ 方程，至少需要纳入一个满足前述工具变量分析条件的外部变量($Z_1$)，这些条件为：(1)它必须对 $X_2$ 有直接的因果作用，

但对 $Y_2$ 没有；(2)它必须是外生的，因而与 $U_1$ 和 $U_2$ 都不相关。同样，至少需要再纳入一个外生的且与 $Y_2$ 有直接因果关系但对 $X_2$ 没有影响的外生变量($Z_2$)才能使 $X_2$ 方程可识别。可以想象，这样的变量很难找到，尤其是在进行跟踪数据二次分析时，因为在调查中并未特意测量此类工具变量。另外，这些外部变量需要与内生变量具有相当强的关联，才能提高第二阶段估计的效率。研究者不能仅仅为了模型的识别而任意给外部变量施加限制，因为所得估计结果的偏差程度与工具变量违反必要假定的程度直接相关。经过理论斟酌而推导出来的工具变量是最理想的。在其他情况下，逻辑理由能够支持某些关系的设定，如 Jennings 和 Niemi 就在一个交互效应模型中，将父母的政党认同设定为每对夫妇自身政党认同的工具变量(Jennings & Niemi, 1975)。在理论上，很难假定妻子父母的政党认同会对丈夫的政党认同产生直接的因果影响，反之亦然。

在实践中，使用个体层次数据进行面板分析时，通常使用社会背景变量，如受教育水平或年龄作为工具变量，因为它们常常被认为是固定的外生变量，因而与误差项不相关。但是，如果这些变量在第一阶段方程中对 $X_2$ 或 $Y_2$ 的影响不大，那么第二阶段方程中所有系数的标准误都将变得相对不精确。另外，如果作为 $Y_2$ 工具变量的 $Z$ 变量事实上对 $X_2$ 有直接影响，那么第二阶段的方程将会设定错误，所得结果将是有偏的参数估计。就此而言，最理想的情况是跟踪调查之初就已将工具变量作为调查设计的一部分，因而能事先设定一些对 $X_2$ 或 $Y_2$ 影响相对较大，但与它们对应方程中的误差项无关的外生变量。

带有合适工具变量的交叉滞后作用及同步作用共存模型可以通过两阶段最小二乘法或在 LISREL 框架下用最大似然法进行估计。TSLS 的估计过程如前所述：要得到 $Y_2$ 方程的系数，就要将 $X_2$ 对 $X_1$、$Y_1$ 及 $Z_1$ 进行回归产生 $\hat{X}_2$，然后将 $Y_2$ 对 $X_1$、$Y_1$ 及 $\hat{X}_2$ 进行回归；要得到 $X_2$ 方程的系数，就要将 $Y_2$ 对 $Y_1$、$X_1$ 及 $Z_2$ 进行回归产生 $\hat{Y}_2$，然后将 $X_2$ 对 $Y_1$、$X_1$ 及 $\hat{Y}_2$ 进行回归。与前面的例子一样，通常建议使用现有统计软件中的 TSLS 程序以正确估计结构系数及模型的有关统计指标。

对 LISREL 矩阵进行简单设置，就可得到(模型的)最大似然估计。将 $X_1$、$Y_1$、$Z_1$ 及 $Z_2$ 设定为外生 $\xi$ 变量，$X_2$ 和 $Y_2$ 设定为内生变量 $\eta_1$ 和 $\eta_2$。来自时间 1

的变量与 $Z$ 变量的效应在连接外生变量和内生变量的 $\mathbf{\Gamma}$ 矩阵中进行设定，设定时，根据工具变量在因果系统中的假定角色，将单个的 $\gamma$ 元素限定为 0。在本例中，从 $Z_1$ 到 $Y_2$ 和 $Z_2$ 到 $X_2$ 的 $\gamma$ 效应都被设定为 0。为了完成因果设置，还需要将 $\mathbf{B}$ 对角线两边的元素都设定为自由状态，以代表时间 2 上 $X$ 和 $Y$ 之间的交互影响。由于总共有 21 个被观察的方差及协方差值，但只有 20 个自由参数(外生变量方差协方差矩阵 $\mathbf{\Phi}$ 的 10 个元素、2 个 $\beta$、6 个 $\gamma$ 以及时间 2 上变量的 2 个 $\psi$ 的方差)，模型仍然还剩一个自由度。如前文所述，这个自由度可对结构扰动项 $U_1$ 和 $U_2$ 之间的协方差等于 0 这一假定进行检验。

估计此类模型时的一个潜在问题是交叉滞后作用及同步作用之间的多元共线性问题。由于受工具变量、因果作用的大小以及变量历时稳定性程度的影响，模型可能无法同时获得对交叉滞后作用和同步因果作用的精确估计。如果共线性很严重，基于对标准误和参数估计间的相关程度的检查，只能对两个因变量对某个因变量的联合效应进行检验。例如，一个同时将群体成员资格对抗议倾向的交叉滞后和同步作用限定为 0 的模型嵌套在同时包含这两个效应的模型内。如果是用二阶段最小二乘法来估计参数，那么可以进行一个 $F$ 检验以确定相对于限制模型，非限制模型是否显著具有更多的解释力。而在LISREL中，$\chi^2$ 值之差在一定的模型自由度之差下，能检验非限制模型是否比不带因果作用的限制模型更能解释观察数据。这样，研究者就能确定一个变量对另一个变量的交叉滞后及同步作用是否联合显著。Mayer 和 Carroll(1987)列出了另外一些 $\chi^2$ 差检验以帮助确定除变量的稳定性外，变量间的所有交叉滞后作用、全部同步作用或者全部因果作用是否统计显著。

上面列出的方法概括了估计二期数据交互效应模型的可能方法。当存在很强的理论依据以假定某一时滞结构时，无论是将同步作用设定为 0，还是将交叉滞后作用设置为 0，模型都可被识别。当研究者事先没有足够的理由将其中一类效应排除在外时，那么带有两种效应的模型只有在纳入外部变量并对它们的因果作用施加了限制之后，才可被识别和估计。不过，如果模型无法拟合数据，放宽这些限定将导致模型无法识别；另外，如果怀疑存在自相关误差，为了进行估计，还必须有额外的工具变量。因此，在很多情况下，两轮数据分析只能

提供非结论性的结果，从而使得三期或多期数据的分析成为必要。

## 三期数据模型

如果有三期或多期数据，估计带有交叉滞后作用及同步作用的模型时，就能更自由地对因果作用参数进行限制或检验。一种可能的做法是，在所关心变量的第二轮及第三轮取值之间设置交叉滞后和同步作用，而将第一轮调查值作为上文 TSLS 或最大似然估计过程中的工具变量。不过，一旦控制 $X_2$ 后，$X_1$ 对 $X_3$（或者控制 $Y_2$ 后，$Y_1$ 对 $Y_3$）没有直接影响，工具变量法就会崩溃。另外，模型将无法提供精确的历时因果过程描述，因为它忽略了第一轮变量和第二轮变量间所有的因果关系。

另一个选择是，通过扩展图 3.2 所示的 LISREL 交叉滞后模型并给模型参数施加其他限制，对三期数据模型进行估计。我们可以在图 3.2 的模型上增加下面这些参数代表全部交叉滞后和同步作用设定：$\beta_9$、$\beta_{11}$ 代表时间 2 和时间 3 上，一个变量对另一个变量的同步作用；$\beta_{10}$、$\beta_{12}$ 代表时间 2 和时间 3 上对应的反向效应。整个模型具有 2 个自由度，因为 6 个观察变量总共有 21 个方差协方差值，而需要估计的自由参数为 19 个——12 个 $\beta$ 参数、6 个结构性项的方差、1 个时间 1 上结构扰动项之间的协方差。不过，第二轮及第三轮内生变量的方程没有一个可解，因为每个预测方程中的内生变量都比遗漏的先定变量多。通过纳入外部先定变量，并对它们的因果作用施加某些限制，是使这些方程可解的一种策略。不过，正如我们已经知道的，这样的变量很可能无法实现，或者此种程序背后所做的假定在给定的情境下站不住脚。识别这一模型的另一可供选择的策略是给参数施加一些相等性限制（即令某些参数相等）。Kessler 和 Greenberg(1981)表明，只需施加下述限定中的两个，即可使模型得到识别，如果将 3 个限定都加上，模型就会过度识别：$\beta_1 = \beta_5$，$\beta_2 = \beta_6$，$\beta_9 = \beta_{11}$。第一个限定使一个变量的稳定性在相邻几轮调查之中相等，第二个限定使第一轮调查与第二轮调查之间的一个变量对另一个变量的交叉滞后作用等于第二轮和第三轮之间对应的作用。研究者应该注意的是，必须而且只需对一个变量的决定因素施加限制。在上述模型中，如果相等性限制被施加在方程 $\eta_5$、$\eta_6$ 的参数上，那么方

程 $\eta_2$、$\eta_3$ 的参数仍然可以不加限制地得到估计。不过，在多数情况下，很少有理论支持我们限制一个变量对另一个变量的效应。研究者可以在初步模型中，对两个变量的效应都施加相等性限定，然后在后续模型中视情况放宽这些限定。如果能够获得多轮调查数据，额外的自由度就能够为我们提供足够的信息，以对模型的参数(包括扰动项的协方差)进行估计，而无需像三期数据时那样设置众多的相等性限制。

我们用来自全国青少年调查的数据检验这个模型，该调查从 1976 年开始，对 1725 名 12 岁至 17 岁的美国青少年进行追踪调查。在此后 4 年中对他们进行了访问，不过这里的分析只限于整个调查的前 3 年(1976 年至 1978 年)。调查测量了青少年自己报告的违法行为、药物及酒精服用情况，还有众多在心理学及社会学文献中被认为可以用来解释青少年违法行为和药物滥用的变量。我们假设个体自报的违法行为在一定程度上由其与违法伙伴的牵涉程度/亲密程度所决定，同样，个体自身的违法行为也会促使其与具有相同行为倾向的同龄人交往(有关 NYS 调查的描述请见 Elliott, Huizinga & Ageton, 1985; Menard & Elliott, 1990)。

我们通过将受访者在 6 种违法行为上的情况取平均值，建构出每个青少年在 1976 年、1977 年、1978 年的一般自报违法量表，我们还询问了过去一年中他们参与所列 6 种行为的频率(取值从代表"从未参与"的"0"，一直到代表"每天 2 到 3 次"的"6")。而每轮调查中的"与违法伙伴的牵涉程度"量表是根据 Elliott 等人(Elliott et al., 1985)提出的方法建构出来的，即将一个测量受访者每天下午、晚上及每个周末与朋友在一起相处的时间长度指标("0"代表"从不"，"5"代表"大量")与另一个测量他们主观感知其朋友中有多少参与了上述 6 种不同违法行为("1"代表"没有人"到"5"代表"全部")的指标相乘。

表 3.2 给出了这一模型几个变形的参数的最大似然估计结果。$\eta_1$ 到 $\eta_3$ 代表 1976 年、1977 年、1979 年与违法伙伴的牵涉程度(IDP)；$\eta_4$、$\eta_5$、$\eta_6$ 代表 3 轮调查中受访者自报的违法情况(SRD)。表格的第一列给出的是对 SRD 及 IDP 的因果作用施加了相等性限定的模型的估计结果，也就是说，该模型设定了如下相等条件：对于 SRD 的方程，$\beta_1 = \beta_5$，$\beta_2 = \beta_6$，$\beta_9 = \beta_{11}$；对于 IDP 的方程，

$\beta_4 = \beta_8$，$\beta_3 = \beta_7$，$\beta_{10} = \beta_{12}$。

**表 3.2 三期数据交互作用模型：自报违法行为及与违法同龄伙伴的交往深度(1976—1978 年)**

| | 模型(1) | | 模型(2) | | 模型(3) | |
|---|---|---|---|---|---|---|
| | 系数 | 标准化系数 | 系数 | 标准化系数 | 系数 | 标准化系数 |
| **稳定效应** | | | | | | |
| SRD | | | | | | |
| $\beta_2$ | 0.40[a] | 0.41 | 0.37[a] | 0.37 | 0.35 | 0.36 |
| $\beta_6$ | 0.40[a] | 0.38 | 0.37[a] | 0.35 | 0.40 | 0.37 |
| IDP | | | | | | |
| $\beta_4$ | 0.54[b] | 0.45 | 0.59[b] | 0.49 | 0.61 | 0.49 |
| $\beta_8$ | 0.54[b] | 0.48 | 0.59[b] | 0.51 | 0.56 | 0.51 |
| **交叉滞后作用** | | | | | | |
| SRD 到 IDP | | | | | | |
| $\beta_3$ | 2.83[c] | 0.37 | 1.74[c] | 0.23 | 1.67 | 0.22 |
| $\beta_7$ | 2.83[c] | 0.42 | 1.74[c] | 0.23 | 1.75 | 0.26 |
| IDP 到 SRD | | | | | | |
| $\beta_1$ | −0.01[d]* | −0.05 | —— | | —— | |
| $\beta_5$ | −0.01[d]* | −0.05 | —— | | —— | |
| **同步作用** | | | | | | |
| SRD 到 IDP | | | | | | |
| $\beta_{10}$ | −1.58[e]* | −0.25 | —— | | —— | |
| $\beta_{12}$ | −1.58[e]* | −0.28 | —— | | —— | |
| IDP 到 SRD | | | | | | |
| $\beta_9$ | 0.09[f] | 0.54 | 0.06[d] | 0.38 | 0.06 | 0.39 |
| $\beta_{11}$ | 0.09[f] | 0.50 | 0.06[d] | 0.35 | 0.07 | 0.35 |
| $R^2(\eta_2)$ IDP(第二轮) | 0.11 | | 0.28 | | 0.26 | |
| $R^2(\eta_5)$ IDP(第二轮) | 0.53 | | 0.54 | | 0.55 | |
| $R^2(\eta_3)$ SRD(第三轮) | 0.13 | | 0.30 | | 0.32 | |
| $R^2(\eta_6)$ SRD(第三轮) | 0.54 | | 0.56 | | 0.55 | |
| $\chi^2$(df) | 167.0(8) | | 168.3(10) | | 163.3(6) | |

注：表中的系数为非标准化估计，标准化系数为带下划线的数字。除带有星号的系数外，所有系数都在 0.05 水平上显著。

a、b、c、d、e、f：系数被限定相等。

资料来源：(美国)全国青少年调查，1976 年、1977 年、1978 年。

这一模型有 8 个自由度，从 $\chi^2$ 值来看，它对数据的拟合非常差。不过，该模型的实质性解释意味着因果交互作用遵循着某种被称为“反馈”的过程(Plewis，1985)，在此过程中，变量间在不同时间间隔上构成相互影响。在这里，变量 IDP 到 SRD 间存在显著的同步作用，而 SRD 到 IDP 之间存在显著的滞后作用。在本例中，这一模式实际上具有合理性，与违法同龄伙伴的交往会影响个体同期的违法行为，而违法行为会使个体在接下来的一年内与具有类似行为倾向的伙伴交往。

第二列给出了移除不显著因果作用后的模型的重新估计情况。这个模型(2)嵌套在模型(1)中，因为除模型(2)中的 $\beta_1$、$\beta_5$、$\beta_{11}$、$\beta_{12}$ 被限定为 0 以外两者具有相同的参数矩阵。因而我们可以使用 $\chi^2$ 值之差来检验模型(1)是否比模型(2)拟合得更好。在这里，$\chi^2$ 值之差为 1.3，在模型之间的自由度之差为 2 的情况下，这是不显著的，因而相对于模型(1)，我们更支持模型(2)。实质上，这些作用表明，从 IDP 到 SRD 的同步作用比 SRD 到 IDP 的滞后作用大 50%，两套作用都处于中等水平。

因为不带同步交互效应的模型已经完全递归了，我们可以放宽前面为了识别模型而施加的相等性限制，并通过模型 $\chi^2$ 值的比较来检验这些限定。表 3.2 的第三列给出了这一估计结果，它表明在放宽相等性限定后，模型拟合水平仅仅提高了一点，估计所得系数的大小也未发生显著的变化。但由于 3 个模型对数据的拟合都不好，模型还需要被进一步调整。

上述模型的估计为我们演示了通过对模型的参数施加相等性限制，以估计复杂多期数据方程组的方法以及 LISREL 和相关程序在估计各种因果作用模型时的灵活性。施加在三期模型参数上的限制比和外部外生变量相关的限制更合理，尤其是当工具变量(最初)并未被作为研究设计的一部分时。

与本章讨论的面板模型相关的一个最后的警告是，估计所得参数应该能表明，该因果系统具有第二章所述意义上的“稳定性”，也就是说，该系统在未来某个时期可以达到平衡。如前文已经提到的，有的模型意味着系统会在某一点上平衡，有的其某些或全部的方差或协方差会无限膨胀而“爆炸”。尽管非稳定系

统从技术上讲并非不可能，但如果所获得的参数估计无法使系统最终达到平衡，那通常被认为是模型设定错误的征兆之一（Arminger，1987）。例如，试想这样一个简单的非递归模型：$Y_t$ 和 $X_t$ 之间交互相关，$\beta_{XY}$ 为 $X$ 对 $Y$ 的影响，$\beta_{YX}$ 为 $Y$ 对 $X$ 的影响，$U_Y$ 为 $Y$ 的误差项，$U_X$ 为 $X$ 的误差项。在这样一个模型中，每个变量都对其自身有间接影响，这样 $Y$ 的变化会通过 $Y$ 对 $X$ 以及 $X$ 对 $Y$ 的影响，最终对自身形成反馈作用。将 $X$ 方程代入 $Y$ 方程就能清楚看到这一点：

$$Y_t = \beta_{YX}\beta_{XY}Y_t + \beta_{YX}U_X + U_Y \quad [3.9]$$

如果 $\beta_{YX}\beta_{XY}$ 大于 1，那么 $Y$ 的增加将使 $Y$ 的方差处于无限增长的过程中。只有当 $\beta_{YX}\beta_{XY}$ 小于 1 时，模型才会稳定或者平衡，因此出现其他参数估计都值得怀疑。在更复杂的模型中，要确定一个模型是否能达到平衡更困难，需要对内生变量之间的相互影响矩阵 **B** 中进行运算。LISREL 可以进行这一检验，名为"稳定性指标"，当其取值小于 1 时，系统可以达到平衡。[10]

# 第 4 章 | 测量误差模型

到目前为止，本书所展示的面板模型都假定分析中所用变量的测量没有误差。不过，现有的测量工具，如态度调查或已出版的汇总统计，对社会科学中多数有意思的概念的测量都是不完美的，这早已得到了很好的证明。既然观察变量含有测量误差，那么，回归模型中结构系数的估计就有可能是错误的，因而研究者可能作出不正确的因果推断。测量误差在面板模型中的问题尤为严重，因为测量误差可使本来没有发生变化的变量看起来发生了变化。不过，追踪研究设计中对相同变量的重复测量能够提供有用的信息，以帮助我们推断观察项目中误差的分量，对剔除了不完美测量的污染效应后自变量的因果作用进行估计。

不过，跟踪调查数据并非解决测量误差问题的万灵丹。所有被考虑的模型只有在其假定条件合理的前提下才是有效的，在很多情况下，这些假定都非常严格。只有在指标的数量或者观察轮次增加的情况下，它们才可被放宽；而在有些情况下，利用上一章讨论过的相等性限定，可以在 LISREL 或相关方法中对这些假定本身进行检验。

## 基本概念

已有不少著作对变量测量误差的一般问题及其后果进行了详细的讨论（Berry & Feldman, 1985; Carmines & Zeller, 1979; Sullivan & Feldman, 1979）。经验研究中使用的绝大多数指标都无法完美地测量所关心的概念，而多少含有一些误差。这些误差或是系统的，或是随机的。在前一种情况下，测

量工具被认为测量了所关心的理论变量之外的其他因素；在后一种情况下，指标测量的是真实概念加上一些“噪音”成分。例如，在抽样调查数据中，随机噪音会因含糊的提问、有限的回应选项、猜想及其他在特定时点上影响受访者言语回应的随机因素而出现。系统误差增加时，指标的效度会不断下降，而随机误差的增加则与指标信度的下降相关。此处讨论的模型将集中关注经验指标的随机误差的校正。有关探测和处理系统误差的更多具体内容，请看 Carmines 和 Zeller(1979)的著作。

$y_{kt}$——真实得分或称“潜变量”$\eta_t$ 在时间 $t$ 的第 $k$ 个指标——的基本随机测量误差模型可以这样表示：

$$y_{kt} = \lambda_{kt}\eta_t + \varepsilon_{kt} \qquad [4.1]$$

其中，$\lambda_{kt}$ 为将 $y_{kt}$ 与潜变量 $\eta_t$ 连接起来的非标准化系数，$\varepsilon_{kt}$ 为随机误差项。我们使用 LISREL 符号系统来表示变量及测量参数，以方便在本章使用这些程序估计有关模型。在单指标测量模型中，$\lambda_{kt}$ 被设定为 1，以使指标的尺度与潜变量的相等。在多指标模型中，必须将一个 $\lambda_{kt}$ 设定为 1，以对潜变量的尺度进行设定。[11]对于 $\varepsilon_{kt}$，通常假定其均值为 0，方差在 $\eta_t$ 的各个水平上都相等，且与 $\eta_t$ 的结构扰动项不相关。在单指标情况下，如果满足这些条件，观察指标的方差 $\sigma^2_{y_{kt}}$ 将等于：

$$\sigma^2_{y_{kt}} = \sigma^2_{\eta_t} + \sigma^2_{\varepsilon_{kt}} \qquad [4.2]$$

其中，$\sigma^2_{\eta_t}$ 为潜变量 $\eta_t$ 的方差，$\sigma^2_{\varepsilon_{kt}}$ 为 $y_{kt}$ 随机误差成分 $\varepsilon_{kt}$ 的方差。因此，观察指标的方差由两个部分组成：“真实得分”方差和随机误差方差。前者由潜变量 $\eta_t$ 的方差代表。一个观察指标的信度 $\rho_{yy}$ 就是“真实得分”方差在指标方差中所占的比例，即：

$$\rho_{yy} = \frac{\sigma^2_{\eta_t}}{\sigma^2_{\eta_t} + \sigma^2_{\varepsilon_{kt}}} \qquad [4.3]$$

换言之，信度是真实得分方差与观察指标方差之比，其值越接近于 1，该潜在概念的指标所含的随机误差就越小。

如果 $\lambda_{kt}$ 未被设定为 1(多指标及标准化测量模型中就是这样)，那么一个指

标的信度为：

$$\rho_{yy}=\frac{\lambda_{kt}^{2}\sigma_{\eta}^{2}}{\lambda_{kt}^{2}\sigma_{\eta}^{2}+\sigma_{\varepsilon_{kt}}^{2}} \qquad [4.4]$$

如果带有随机测量误差的指标被用于回归方程的自变量，那么所得结果将是潜变量真实因果作用的有偏估计，因为可以证明自变量不再与估计方程的扰动项无关(Markus, 1979:55)。[12]在双变量回归模型中，估计系数等于真实系数乘以指标的信度，因而信度差的测量将导致对变量效应的低估或使其变小。但在多变量模型中，偏差的方向可能为任意方向，且任意一个自变量存在测量误差，都可能造成模型中所有其他变量的效应的估计严重有偏，即便那些变量本身被假定没有测量误差。

从某些方面来看，测量误差问题在跟踪调查情境下甚至更严重。正如本书已经表明的那样，绝大多数面板模型都是条件变化模型或稳定得分模型的变种，其中，$Y_t$ 是 $Y_{t-1}$ 及某些自变量 $X_t$、$X_{t-1}$ 等的函数。在第 2 章中我们已经注意到，滞后内生变量 $Y_{t-1}$ 在这些模型中充当了几个角色：作为系统向均衡点运动时，“向均值回归”及“负反馈效应”的控制项；作为被遗漏因果路径的代理之一；影响 $Y_t$ 的一个实际变量。遗漏 $Y_{t-1}$ 通常会导致参数估计偏差，因为 $Y_{t-1}$ 和 $\Delta Y$ 之间很可能存在负相关，但如果 $Y_{t-1}$ 含有如方程 4.1 所示的测量误差，而且 $\lambda_{kt}=1$，那么 $\varepsilon_{kt}$ 之间互不相关且与真实得分 $\eta_t$ 无关。这是可以证明的，只要误差方差在各个时点相等：

$$\text{Cov}(Y_{t-1}, \Delta Y)=\text{Cov}(\eta_{t-1}, \Delta\eta)-\text{Var}(\varepsilon_{t-1}) \qquad [4.5]$$

也就是说，观察的 $Y_{t-1}$ 及其后续变化间的协方差等于 $t-1$ 时的“真实得分”与“真实后续变化”之间的协方差减去 $Y_{t-1}$ 的测量误差方差。这个方程意味着，除非 $Y_{t-1}$ 中的测量误差被控制，否则迄今所分析的条件变化模型都将导致 $Y_{t-1}$ 与其后续变化之间的协方差过于负向的估计，低估 $Y_{t-1}$ 对 $Y_t$ 的作用(相对于“真实得分”$\eta_{t-1}$ 对 $\eta_t$ 的作用)。这反过来通常会导致高估其他变量对 $Y_t$ 的因果作用，因为它们与观察的 $Y_{t-1}$ 正向相关。相反，如果 $Y$ 的测量误差在各个时点间相关且未被考虑，这将导致对 $Y_{t-1}$ 对 $Y_t$ 作用的高估，并相应地低估其他变

量的影响。因此,滞后外生变量的测量误差是妨碍在面板模型下成功进行因果推论的最严重的障碍之一,因而在估计结构效应时,建立将测量误差纳入的模型非常重要。

尽管测量误差在面板分析中会带来严重的问题,但仍比截面数据情况下处理起来容易得多。在截面数据中,有两种程序通常被用来处理测量误差的问题。一是工具变量法或者二阶段最小二乘估计,与上一章为结构方程模型讨论过的程序相似。由于存在测量误差时,OLS 估计中的偏差是由自变量与方程误差项之间的相关造成的,所以一个可能的解决方法是在模型中纳入一个与真实得分或潜变量相关,但与其随机误差项无关的外生变量(在定义上,就与结构性扰动项无关)(Berry & Feldman, 1985:34—37)。

对于面板数据,还存在其他处理测量误差的策略,而且多数情况下比工具变量法或二阶段最小二乘法好。由于找到合适的工具变量很难,尤其是在二次分析中,二阶段法依赖于在特定模型中可能并不成立的关于工具变量与扰动项之间无关的假定,且可能带有误差的自变量与其工具变量之间的低相关常常会导致第二阶段的估计很不精确。另外,尽管工具变量法在理想情况下能够为将模型中的潜变量连接起来的结构参数提供一致的估计,但它也无法提供与指标的信度、误差方差或其他测量特性等相关的直接信息,而这些往往是经验研究的重要目标。因此,工具变量法给面板分析带来了一些不足。

另一个发展出来以应对测量误差的方法是多指标法,这种方法使用同一个潜变量的多个测量,同时产生对特定模型中的结构效应和测量参数的估计(Sullivan & Feldman, 1979)。在跟踪调查设计中,于不同时间对同一指标的重复测量大大增强了这一方法的效力,因为额外轮次的数据为有关结构系数及测量系数的估计提供了更多信息。事实上,只要至少有 3 轮数据,就能估计出每个潜变量只有一个指标的模型的测量特性及结构效应,因而所有多期面板模型都可被当做多指标法的变种。但是,这种方法的全部效力体现在多指标、多轮测量模型中,在有些情况下,它甚至可能估计不同时点的测量误差之间及结构性扰动项之间相关的模型。这样的分析能够提供每个指标项的信度、潜变量的

稳定性以及因果模型中，将潜变量连接起来的结构效应等方面的最全面的信息。

## 单指标模型

### 二期模型

当只有二期数据可得时，用来处理单指标 $y_t$ 的测量误差的选择有限。为便利起见，这里遵循的是 LISREL 类型的标记法，如果令 $y_t$ 为潜变量 $\eta_t$ 在时间 $t$ 上的观察指标，$\varepsilon_t$ 为 $y$ 在时间 $t$ 上的误差方差，$\beta_1$ 为将 $t-1$ 时刻与 $t$ 时刻的潜变量联系起来的稳定性系数，$\lambda_k$ 为外生变量 $\xi$ 对 $\eta_t$ 的效应，$\zeta_t$ 为潜变量的结构扰动项，我们可以将关心的结构模型表达为：

$$\eta_t = \beta_1 \eta_{t-1} + \sum \gamma_k \xi_k + \zeta_t \qquad [4.6]$$

不过，使用并非完美测量的观察变量 $y_{t-1}$ 和 $y_t$ 意味着，估计模型不仅含有真实得分 $\eta$，也含有测量误差：

$$(\eta_t + \varepsilon_t) = \beta_1(\eta_{t-1} + \varepsilon_{t-1}) + \sum \gamma_k \xi_k + \zeta_t \qquad [4.7]$$

从这一方程中，我们可以清楚地看到模型无法识别，因为即便我们假定两个指标的测量误差的方差相等，也无法分离观察变量中真实得分和误差各自所占的比例。

McAdams(1986)指出了几种通过给方程中的参数施加额外限制，以便在 LISREL 中估计这一模型的备选方法。分析者可以将 $\beta_1$ 的取值限定为 1，这样，模型就被简化为第 2 章中讨论的无条件变化—得分模型的一个变种，其中 $\Delta y$ 由一系列 $X$ 变量进行预测。另一种可能的方法是限定 $\zeta_t$ 的取值，从而设定方程中总共被解释掉的方差。最后，可以限定 $\varepsilon_t$，将 $y_t$ 的信度设定为某个已知值或假定值。

对于二期数据，这些方法可被当做解释工具，用来产生对方程 4.7 中参数的

一系列估计，而且通过评估这些结果，分析者可以获得从对模型的不同限定中呈现出来的对潜在因果过程的描述，或得到方程参数的可能取值范围的估计。但识别模型所必需的限定存在一些问题。将 $\beta_1$ 限定为 1 忽略了所有来自于 $y_t$ 而非测量误差对 $\Delta y$ 的影响，而限定 $\zeta_t$ 或 $\varepsilon_t$ 的方差常常也是站不住脚的，因为我们并不能事先知道方程的总体误差方差或观察指标的误差方差分量的可能值。显然，我们希望能在没有这些限定的情况下对 $\beta_1$ 值以及 $\zeta_t$ 和 $\varepsilon_t$ 的方差进行估计，而这使更多轮次的信息成为必要。

## 三期模型

分析一个潜变量只有一个指标的模型，被 Wiley 等人(1970)以及 Heise (1969)发展并用于三期数据。三期模型一般以图表方式呈现，如图 4.1 所示。

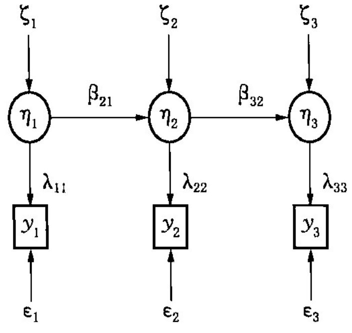



**图 4.1　三期数据单指标测量模型**

将指标和它们的潜变量联系起来的测量方程可以写为：

$$\begin{aligned} y_1 &= \lambda_{11}\eta_1 + \varepsilon_1 \\ y_2 &= \lambda_{22}\eta_2 + \varepsilon_2 \\ y_3 &= \lambda_{33}\eta_3 + \varepsilon_3 \end{aligned} \qquad [4.8]$$

其中，$\lambda_{tt}$ 为连接每个指标 $y$ 与其对应潜变量 $\eta$ 的系数。刻画潜变量(内生变量)之间的因果联系的结构方程为：

$$\eta_1 = \zeta_1$$

$$\eta_2 = \beta_{21}\eta_1 + \zeta_1 \qquad [4.9]$$

$$\eta_3 = \beta_{32}\eta_2 + \zeta_3$$

对于测量误差和结构扰动项的一套标准假定是，$\varepsilon_t$ 与 $\eta_t$ 之间、$\varepsilon_t$ 与 $\zeta_t$ 之间的协方差都等于 0。另外，测量误差被假定服从均值为 0、方差稳定且是不同时点间彼此互不相关的随机分布。扰动项 $\zeta_t$ 同样被假定是随机的、均值为 0、方差稳定且不同时点间互不相关。

现在这一模型的参数尚不能被识别。因为有 11 个"未知"或自由参数(3 个 $\varepsilon_t$、3 个 $\lambda_{tt}$、3 个 $\zeta_t$ 方差以及 2 个 $\beta$ 系数)，但只有 6 个"已知的"观察变量的方差和协方差值，没有足够的信息用以获得对这些参数的唯一估计。因此，我们还需要作出额外的假定，向模型参数施加某些限定。

在 Heise 法中，模型的识别是通过同时标准化潜变量和观察指标，并限制指标 Y 的信度在 3 个时期相等的方式达到的。通过将模型标准化，$\eta_t$ 的方差等于 1，这意味着模型结构部分的自由参数只有稳定性系数 $\beta_{21}$ 和 $\beta_{32}$。因为观察到的指标同样被标准化了，模型测量部分仅有的自由参数是 $\lambda_{tt}$，因为 $\varepsilon_t$ 的方差必定等于 $1-\lambda_{tt}^2$。[13]根据方程 4.4 可以推断出指标的信度就是 $\lambda_{tt}^2$，因为 $\eta_t$ 的方差和方程 4.4 的分母都等于 1。使信度在各轮调查之间相等，亦即 $\lambda_{tt}$ 都相等，得到的是一个恰好可识别的模型，其中有 3 个自由参数和 3 个用以估计它们的观察指标间的相关系数。就观察到的相关系数而言，解出这 3 个参数就很简单：

$$\lambda = \sqrt{\frac{r_{12}r_{23}}{r_{13}}}$$

$$\beta_{21} = \frac{r_{13}}{r_{23}} \qquad [4.10]$$

$$\beta_{32} = \frac{r_{13}}{r_{12}}$$

$y$ 的信度就是方程 4.10 第一个表达式的平方，因为 $\lambda^2$ 等于 $y$ 的方差中被潜变量 $\eta_t$"解释掉"的部分。

Wiley-Wiley 法是通过限定指标的误差方差在各时期相等，也就是说 $\sigma^2_{\varepsilon_1}=\sigma^2_{\varepsilon_2}=\sigma^2_{\varepsilon_3}$，在此过程中得到两个自由度，使模型在非标准化形式下得到识别。在非标准化模型中，由于 $\eta_t$ 的方差不再被限定为 1，所以不可观察的 $\eta_t$ 变量的刻度尚未被确定的问题是通过使每个 $\lambda_{it}$ 都等于 1 来解决的。在这些假定条件下，以如下方式对观察变量的方差及协方差进行代数运算，即可得到各参数的估计。将 3 个潜变量方程 4.9 的结构方程预测值代入测量方程 4.8，得到如下结果：

$$
\begin{aligned}
y_1 &= \zeta_1 + \varepsilon \\
y_2 &= \beta_{21}\zeta_1 + \zeta_2 + \varepsilon \\
y_3 &= \beta_{32}(\beta_{21}\zeta_1 + \zeta_2) + \zeta_3 + \varepsilon
\end{aligned}
\tag{4.11}
$$

取这些 $y$ 的方差和协方差的期望值，得到：

$$
\begin{aligned}
\mathrm{Var}(y_1) &= \mathrm{Var}(\zeta_1) + \mathrm{Var}(\varepsilon) \\
\mathrm{Cov}(y_1y_2) &= \beta_{21}\mathrm{Var}(\zeta_1) \\
\mathrm{Cov}(y_1y_3) &= \beta_{21}\beta_{32}\mathrm{Var}(\zeta_1) \\
\mathrm{Var}(y_2) &= \beta_{21}^2\mathrm{Var}(\zeta_1) + \mathrm{Var}(\zeta_2) + \mathrm{Var}(\varepsilon) \\
\mathrm{Cov}(y_2y_3) &= \beta_{32}[\beta_{21}^2\mathrm{Var}(\zeta_1) + \mathrm{Var}(\zeta_2)] \\
\mathrm{Var}(y_3) &= \beta_{32}^2[\beta_{21}^2\mathrm{Var}(\zeta_1) + \mathrm{Var}(\zeta_2)] + \mathrm{Var}(\zeta_3) + \mathrm{Var}(\varepsilon)
\end{aligned}
\tag{4.12}
$$

可以看到，模型恰好可识别，带有 6 个自由参数（3 个 $\zeta_t$ 的方差，$\beta_{21}$、$\beta_{32}$ 及 $\varepsilon$ 的方差）和 6 个用来估计这些参数的方差和协方差观察值。求解 $\varepsilon$ 的方差及两个稳定系数，得到：

$$
\begin{aligned}
\mathrm{Var}(\varepsilon) &= \mathrm{Var}(y_2) - \frac{\mathrm{Cov}(y_2y_3)\mathrm{Cov}(y_1y_2)}{\mathrm{Cov}(y_1y_3)} \\
\beta_{21} &= \frac{\mathrm{Cov}(y_1y_2)}{\mathrm{Var}(y_1) - \mathrm{Var}(\varepsilon)} \\
\beta_{32} &= \frac{\mathrm{Cov}(y_1y_3)}{\mathrm{Cov}(y_1y_2)}
\end{aligned}
\tag{4.13}
$$

一旦得到这些估计，根据方程 4.3 计算各个指标的信度就非常简单了。因为在这一正好识别的模型中，$Var(\eta_t)$恰好等于 $Var(y_t) - Var(\varepsilon)$，所以每个指标的信度即 $[Var(y_t) - Var(\varepsilon)]/Var(y_t)$。

## 在 LISREL 中整合单指标测量模型和结构模型

至此，所讨论的模型的系数都可以通过 LISREL 中的最大似然法获得。尽管在这种恰好识别的情况下，所得的参数估计将与上面描述的完全一样，但 LISREL 能够提供最大似然标准误，这用其他方法是很难估计的。而且这一方法很容易被扩展，以估计四期或期次更长的单指标模型，或者潜变量带多个指标的模型的参数。LISREL 还可以用来对上两章描述过的更复杂的结构模型情景下的测量误差进行建模，而且，如果这样的模型过度识别，LISREL 不仅能提供对所有参数的高效估计，还能提供对模型拟合优度的检验。

我们将通过估计第 3 章的政党认同与候选人评价模型的测量及结构参数来说明这些方法。所得结果呈现在表 4.1 中。

**表 4.1　带测量误差的三期模型**

| | 模型(1) | 模型(2) |
|---|---|---|
| **误差方差** | | |
| $\theta_\varepsilon$(政党认同) | 0.55 | 0.55 |
| $\theta_\varepsilon$(候选人评价) | 225.7 | 225.8 |
| **稳定性** | | |
| $PID_1$-$PID_2$ | 1.02 | 1.01[a] |
| | 0.98 | 0.98 |
| $PID_2$-$PID_3$ | 1.01 | 1.01[a] |
| | 1.00 | 1.01 |
| $THERM_1$-$THERM_2$ | 0.98 | 0.85[b] |
| | 0.87 | 0.78 |
| $THERM_2$-$THERM_3$ | 0.90 | 0.85[b] |
| | 0.84 | 0.79 |
| **交叉—滞后作用** | | |
| $PID_1$-$THERM_2$ | —— | 28.2[c] |
| | | 0.15 |
| $PID_2$-$THERM_3$ | —— | 28.2[c] |
| | | 0.15 |

续表

| | 模型(1) | 模型(2) |
|---|---|---|
| $THERM_1$-$PID_2$ | —— | 0.001[d]*** |
| | | <u>0.01</u> |
| $THERM_2$-$PID_3$ | —— | 0.001[d]*** |
| | | <u>0.01</u> |
| **信度** | | |
| $PID_1$ | 0.86 | 0.86 |
| $PID_2$ | 0.87 | 0.87 |
| $PID_3$ | 0.87 | 0.87 |
| $THERM_1$ | 0.81 | 0.81 |
| $THERM_2$ | 0.84 | 0.84 |
| $THERM_3$ | 0.85 | 0.85 |
| $\lambda^2$(df) | 0.0(0) | 12.7(7) |

注:表中的数字为最大似然估计值。带下划线的为标准化系数。不带星号的系数都在统计上显著。
所有模型的 $N$ 都是 733。
a、b、c、d:系数被限定必须相等。
资料来源:美国全国选举研究,主要面板文件,1980 年。

表格第二列包含的是政党认同和候选人评分变量的测量误差及真实得分历时稳定性的分离 Wiley-Wiley 法(separate Wiley-Wiley)估计结果。如果在 LISREL 中设定这些模型,那么连接潜变量的 $\eta_t$ 系数为 **B** 矩阵的非对角元素,代表结构扰动项的系数为 **Ψ** 矩阵的对角元素,而将潜变量及其观察指标连接起来的 $\lambda_y$ 系数是 $\mathbf{\Lambda}_y$ 矩阵的主对角元素,最后测量误差方差为 $\mathbf{\theta}_\varepsilon$ 的主对角元素。$\lambda_y$ 被设定等于 1,而 $\theta_\varepsilon$ 被限定在各轮调查中相等。估计得到政党认同量表的误差方差为 0.55,而政党认同真实得分的标准化稳定性从第一轮到第二轮调查为 0.98,从第二轮到第三轮为 1.0。与观察得到的第一轮、第二轮之间的相关 0.87,第二轮、第三轮之间的相关 0.88 相比,这一估计表明,校正指标的测量误差后,真实得分的稳定性近乎完美。通过方程 4.3 给出的公式计算得到量表的信度,在第一轮中为 0.86,第二轮中为 0.87,第三轮中为 0.87。候选人评价变量的对应结果为,估计的误差方差 225.5,校正的稳定性估计 0.88、0.89,而估

计的信度在第一轮调查中为0.81,第二轮为0.84,第三轮为0.85。

这些方法能以几种方式用于变量间还含有结构效应的模型。方法之一是将指标的误差方差固定为 Wiley-Wiley 初始估计所得值,然后将这一信息加给前面两章估计过的任意一个模型,就能对结构效应进行估计。或者,只要所有参数都可识别,LISREL 就能同时给出对测量误差和结构系数的估计。作为这一方法的例子,表4.1的第三列给出了第3章的三期交叉滞后交互作用全模型的重新估计结果。模型显示,仅政党认同对候选人评分存在显著影响,这和前面假定不存在测量误差的模型估计得到的交互作用形成了对比。政党认同及候选人评价的误差方差估计结果在统计上显著。此外,模型对于数据的拟合非常好,在7个自由度下,$\chi^2$ 为12.5,这意味着该模型在总体中无法被拒绝($p > 0.05$)。上述模型证明,一旦测量误差得到考虑,从面板模型中得到的因果作用可能非常不同,从而引导研究者在变量的实际关系上作出不同的结论。

### 四期或更长的单指标模型

Wiley-Wiley 模型及 Heise 模型在三期面板数据下,对估计指标的测量特性和结构效应非常有用。不过,这些模型中的假定和限制有时很难成立。具体来讲,Heise 模型中信度相等的假定以及 Wiley-Wiley 模型中误差方差相等的假定尤其成问题,因为并没有先验的理由支持潜变量的方差(出现在信度的计算中),或者误差方差将在不同时期保持不变。[14]事实上,Jagodzinski、Kühnel 和 Schmidt(1987)表明,在很多情况下,指标的误差方差会随着时间的推移而逐渐变小,从他们的观点来看,这能够解释为什么“Wiley-Wiley 以及 Heise 提出的模型……常常在调查间隔短的跟踪研究中失败”(1987:294—295)。

另外,由于模型恰好被识别(如果未纳入其他内生或外生变量),因而无法对模型的整体拟合水平进行检验,因为它们都能完美地说明观察数据。在有些情况下,不合情理的经验结果(如在 Wiley-Wiley 法中出现负的误差方差估计,或者在两种方法下信度或标准化稳定性系数的估计明显大于1)是模型不适用的迹象。但无法确定究竟是测量模型设定错误,还是结构模型设定错误。

最后，有关测量误差不相关的假定也是有问题的，因为除了与潜变量有关系外，相同题项可能在不同时期间存在某些共变。调查中出现测量误差相关的原因，包括记忆效应、题项措辞或意义相似而导致的不同时期的回应相似、潜变量（该题项正是其指标）产生的因果影响的独立性等。忽略误差相关的实际后果是极有可能夸大真实得分的稳定性，因为由测量误差相关造成的真实得分稳定性也被合并到了估计的 $\beta$ 参数中。

更多轮次调查信息的加入会使单指标测量模型的估计和评价变得容易一些（实际运用见 Feldman，1989；Green & Palmquist，1990）。其优势之一是，调查轮次更多的模型的某些参数可能会过度识别。在这种情况下，模型整体拟合的检验可以通过利用代数方式分别计算出过度识别参数，然后比较其取值的方式进行，也可以利用 LISREL 获得参数的最大似然估计，并使用 $\chi^2$ 拟合优度检验来评估模型对观察指标的方差和协方差的解释程度。Werts、Jöreskog 和 Linn(1971)扩展了图 4.1 中的 Wiley-Wiley 模型以用于四期数据，并表明“处于中间的”指标 $y_2$ 和 $y_3$ 的误差方差和真实得分方差（进而信度）能够被识别，而从时间 2 到时间 3 的潜变量的稳定性 $\beta_{32}$ 的估计被过度识别。如果用观察的相关系数表示，$\beta_{32}=r_{13}r_{24}/r_{12}r_{34}$ 及 $\beta_{32}=r_{14}r_{23}/r_{12}r_{34}$。如果调查轮次更多，将会有更多的参数被过度识别。

期数更多的面板分析的第二个优势为，额外信息允许对带有相关的误差的模型进行更高效的检验，并大大提高将此类模型与带随机误差的模型进行比较的能力。Wiley-Wiley(1974)提出了针对三期面板的测量误差相关的模型，但这种正好识别的模型常常会得到反常结果，如误差方差和协方差为负、不合理的稳定性估计等。对于四期数据，假定其稳定性相等、误差方差相等及误差协方差相等的模型是带 3 个自由度（10 个观察的方差和协方差减去 7 个自由参数——4 个扰动项方差 $\psi$、1 个 $\beta$、1 个误差方差和 1 个误差协方差）的过度识别模型。不过，Palmquist 和 Green(1989)表明，来自这一模型的估计在指标信度及真实得分稳定性较低的情况下，相对缺乏效率。另外，他们注意到，随着调查轮次的增加，稳定性、误差方差及误差协方差相等的可能性将显著下降。但四期数据仍提供了显著的优势，能通过 $\chi^2$ 拟合优度测量及对参数估计可能性的检

查来对各种替代模型进行评估。

## 多指标模型

当一个潜变量有多个指标可用时，单指标测量模型的很多局限都能被突破。如果存在两个指标，有些限制性模型依靠两轮测量就可以被估计。可得的指标及调查期次越多，包含历时相关测量误差的过度识别模型也就越可能被估计，并可通过整体的拟合优度统计进行检验。研究者也可能给模型的参数施加各种限定，然后通过放宽这些限定并比较各替代模型的概要性 $\chi^2$ 测量的方式来检验这些限定。通过这些方式，多指标法在估计和评估测量模型上，能够提供大得多的灵活性和效力。

### 二期双指标模型

对迄今讨论过的模型的最简单扩展是给潜变量 $\eta_t$ 增加另外一个指标 $y_{2t}$。增加这一指标可以使图 4.2 所示的测量模型仅需二期数据就能被估计出来。

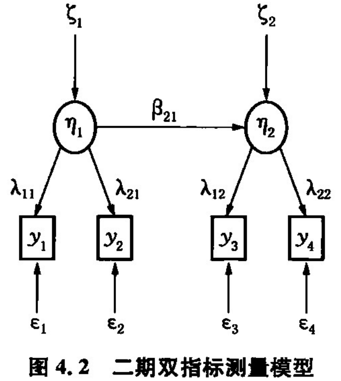



**图 4.2　二期双指标测量模型**

在这一模型中，$Y_1$ 和 $Y_3$ 分别为潜变量 $\eta$ 在时间 1 和时间 2 上的同一个指标，$Y_2$ 和 $Y_4$ 是 $\eta$ 在每轮调查中的第二个指标。滞后潜变量被设定对其第二轮取值存在影响。我们对这一模型作出了如下初始假定：(1)所有变量、误差项及结构扰动项都被表达为离均差形式；(2)误差项 $\varepsilon_t$ 和 $\zeta_t$ 服从正态分布；(3)$\varepsilon_t$、$\eta_t$ 与

$\zeta_t$ 不相关;(4)$\eta_1$ 与 $\zeta_2$ 无关;(5)不同时点的 $\zeta_t$ 互不相关;(6)测量误差项之间互不相关。潜变量尺度的设定是通过将 $Y_1$ 和 $Y_3$ 的系数固定为 1(当然,也可以对 $Y_2$ 和 $Y_4$ 这么操作)完成的。

模型有 9 个自由参数需要估计($\zeta_1$ 和 $\zeta_2$ 的方差、$\beta_{21}$、$\lambda_{21}$、$\lambda_{22}$ 以及 4 个误差方差),观察指标的 10 个观察的方差和协方差使最终还剩 1 个自由度可用对模型进行检验。只要通过 LISREL 或相关方法得到这些估计值,就可如上面的方程 4.4 一样计算出各个指标的信度。

除了只需二期数据就可得到估计外,这一模型还有几个令人满意的特征。第一,研究者可以采取不同的限定方式以检验各指标的测量特性,而且每一个限定都能得到检验其适用性所必需的自由度。在心理测量学中,如果 $\lambda_{11}=\lambda_{22}$ 且 $\theta_{\varepsilon_1}=\theta_{\varepsilon_2}$,那 $Y_1$ 和 $Y_2$ 被称为"$\eta_1$ 的平行测量"。这在 LISREL 中可以通过将两组 λ 系数设定为 1,且令每个时点内的两个误差方差相等的方式进行检验。由此增加的 4 个自由度可以用来检验这些限定的适用性。或者,如果 $\lambda_{11}=\lambda_{22}$,但误差方差互不相等,那么 $Y_1$ 和 $Y_2$ 就是 $\eta_1$ 的 tau 相等测量。这只需要令两个 λ 系数都等于 1,就能得到检验,结果得到 2 个自由度。在图 4.2 的模型中,这种测量被称为"同源的测量",连接指标与潜变量的系数互不相同,指标的误差方差也是如此。

通过令同一个指标在不同时点的误差方差相等(即 $\theta_{\varepsilon_1}=\theta_{\varepsilon_3}$)或同一个指标在不同时间的 λ 系数相等(即 $\lambda_{11}=\lambda_{12}$),还可以对模型施加更多的限定,这些限定都可以用模型相对拟合的方式进行检验。后面这个令同一个指标的 λ 在所有时期相等的限定,在面板模型中尤其重要,这相当于假定一个非标准化潜变量在每轮调查中,对某个指标的非标准化作用相同。Wheaton 等人(1977:129)断言,这一假定是下面这种模型的关键要素,这种模型"标榜要解释(潜)变量的稳定性,这意味着我们指望能(在不同时间)测量同一个概念。严格地讲,稳定性这一概念需要这一设定"。因此,限定 λ 在不同时间相等的假定通常都会出现,对这些假定的检验在 LISREL 或相关方法中通过 $\chi^2$ 差异检验,将对应模型与非限制模型进行比较即可实现。

双指标二期模型的主要不足在于,它没有在不同时间测量的相同题项之间设定相关的误差。Costner(1969)以及 Sullivan 和 Feldman(1979)发现,图 4.2

的模型中唯一一个过度识别的参数是稳定性参数，它有两种估计。因此模型拟合度的评估可通过 LISREL 的 $\chi^2$ 或计算出这两个估计然后比较其取值的方式进行，可以对不同时间的测量误差是否存在相关进行检验，因为（如果误差相关是正的）模型估计的真实得分的稳定性将不足以解释指标在不同轮次间被观察到的相关。遗憾的是，一旦纳入相关的测量误差，图 4.2 中的模型就变得不可识别。要想估计那些参数，有 3 种方法可供选择：(1)在二期模型中，给潜变量增加更多的指标；(2)在二期模型中至少纳入一个在两个时点都会影响潜变量的外生“背景”变量；(3)增加更多轮次的数据信息。所有这些方法都能在不对测量误差矩阵施加任何限制的情况下，使模型得到识别。第一种和第三种方法更可取，因为这样所得的模型将有足够的自由度向参数估计施加或放宽各种限定。第二种方法在额外指标或更多轮次的信息都不可得时，是必然选择。

## 一个三指标二期模型的例子

图 4.3 展示了一个二期模型，其中合法性政治抗议倾向有 3 个指标，所用数据来自前面章节描述过的前联邦德国政治抗议数据。可以看到，同一指标不同时期间的相关误差，由图中同一指标在两个不同时点间自由的误差协方差代表，也就是 $\theta_{\varepsilon_{41}}$、$\theta_{\varepsilon_{52}}$ 及 $\theta_{\varepsilon_{63}}$。$Y_1$ 和 $Y_4$ 为第一轮、第二轮调查中，对参与“被批准的游行示威”的倾向的回应，$Y_2$ 和 $Y_5$ 为“与公民行动组织一起工作”的倾向，$Y_1$ 和 $Y_4$ 为“为请愿收集签名”的倾向。

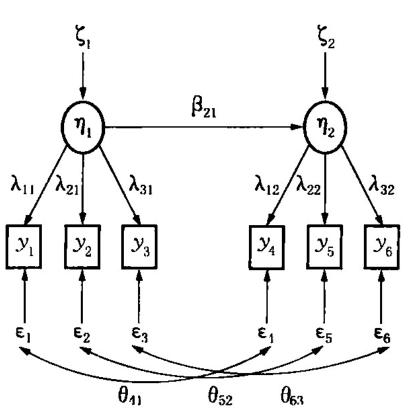



图 4.3　含有相关的测量误差的三指标模型

模型中有(6×7)/2或21个观察的方差和协方差，有16个自由参数——$\zeta_1$ 和 $\zeta_2$ 的方差、$\beta_{21}$、$\lambda_{21}$、$\lambda_{31}$、$\lambda_{22}$、$\lambda_{32}$、6个观察指标的误差方差以及各指标不同时期的误差项之间的3个协方差。潜变量的刻度通过将 $\lambda_{11}$、$\lambda_{12}$ 固定为1而得到设定。模型共有5个自由度。表4.2给出了这一模型的几种变形的最大似然估计。

**表4.2 三指标二期模型：含有和不含相关的测量误差**

| | 相关误差模型(1) | 无相关误差的模型(2) |
|---|---|---|
| 测量系数 | | |
| $\lambda_{11}$ | 1.00[a] | 1.00[a] |
| | 0.68 | 0.70 |
| $\lambda_{21}$ | 1.12* | 1.11* |
| | 0.68 | 0.69 |
| $\lambda_{31}$ | 1.59* | 1.48* |
| | 0.91 | 0.88 |
| $\lambda_{12}$ | 1.00[a] | 1.00[a] |
| | 0.75 | 0.79 |
| $\lambda_{22}$ | 1.11* | 1.06* |
| | 0.77 | 0.79 |
| $\lambda_{32}$ | 1.41* | 1.30* |
| | 0.91 | 0.89 |
| $\theta_{11}$ | 3.92* | 3.66* |
| $\theta_{22}$ | 4.85* | 4.66* |
| $\theta_{33}$ | 1.69* | 2.36* |
| $\theta_{44}$ | 3.34* | 3.11* |
| $\theta_{55}$ | 3.68* | 3.42* |
| $\theta_{66}$ | 1.80* | 2.28* |
| $\theta_{41}$ | 1.47* | —— |
| $\theta_{52}$ | 0.91* | —— |
| $\theta_{63}$ | −0.74* | —— |
| 稳定性系数 | | |
| $\beta_{21}$ | 0.62* | 0.64* |
| | −0.54 | 0.54 |
| $R^2(\eta_2)$ | 0.29 | 0.29 |
| $\chi^2$(df) | 12.5(5) | 69.8(8) |

注：表中的数字为非标准化最大似然估计。标准化的系数带下划线。所有模型的 $N$ 都是377。

a. 系数被固定为1。

* 表示在0.05水平上显著。

表格第二列所示结果对应图 4.3 的相关误差模型。正如该表所示，所有参数估计都在统计上显著，包括各个指标不同时间的测量误差之间的协方差($\theta_{\varepsilon_{41}}$、$\theta_{\varepsilon_{52}}$、$\theta_{\varepsilon_{63}}$)。估计的“签名请愿”指标($Y_3$和 $Y_6$)间的协方差为负，这一不合理的结果将在第 5 章的负自相关背景下讨论。将 $\lambda_y$ 系数取平方即可得到各指标的信度，因为在完全标准化的模型中，$\eta$ 和 $Y$ 的方差都等于 1。[15]在本例中，信度在 0.5 到 0.8 的范围内，表明观察变量中有大量的随机测量误差。潜在“合法抗议倾向”变量的非标准化稳定性被估计为 0.62，标准化作用为 0.54。模型的拟合优度以 $\chi^2$ 测量，为 12.5，带 5 个自由度，表明模型对观察数据的拟合相对较好。

在第三列中，估计了一个不带自由的误差协方差的模型，最终拟合度要差很多，在 8 个自由度下，$\chi^2$ 值为 69.8。由于这一模型除了 $\theta_{\varepsilon_{41}} = \theta_{\varepsilon_{52}} = \theta_{\varepsilon_{63}} = 0$ 的限定之外，与第二列中的模型完全一样，因此，放宽这些限定对模型拟合度的改进可以通过比较两个 $\chi^2$ 而得到检验。在本例中，卡方值之差在 3 个自由度下为 57.3，因此通过允许误差项之间存在协方差而使模型得到的改进在统计上是显著的。不过实际上，误差协方差并不大，因为两个模型的非标准化稳定性相差很小，而在两套标准化的稳定性之间没有差异。

一些其他测量模型同样可以在 LISREL 框架中得到检验。其中一个值得推荐的模型根据 Wheaton 等人(1997)的主张，限定每个指标在第一轮调查和第二轮调查中的 $\lambda$ 系数相等，这可以保证潜变量在不同时间上的相似性。通过施加这一限定，模型得到 2 个自由度，在这种情况下，$\chi^2$ 的差异为 2.7，在 0.05 水平上统计不显著。因而可以得出结论认为这一相等性限制在这一数据中成立。其他的模型还可以限定相似指标在不同时点的误差方差相等，或者如果有理论依据怀疑指标间存在某些与潜变量不相关的共同变异，也可以设定让同一时点的不同指标间的误差协方差自由变化。还可以就总体中的不同子群体比较各个测量系数和潜变量的稳定性。关于群体差异的讨论，可以参阅 Jöreskog 和 Sörbom 的著作(1989：第 9 章)，该方法的应用可以在 Judd、Krosnick 和 Milburn(1981)以及 Porst、Schmidt 和 Zeifang(1987)的著作中

找到。

### 整合多指标模型和结构模型

这种类型的测量模型可以被直接整合进更复杂的因果系统。我们可以设置一些模型，假定图 4.3 中的潜变量被一系列外生变量或内生变量所影响，这些新增变量可以带有误差项，也可以不带。在 LISREL 中，可以对 $\mathbf{B}$、$\mathbf{\Psi}$、$\mathbf{\Lambda}_y$、$\mathbf{\theta}_\epsilon$ 矩阵进行调整，以刻画新的内生变量的测量与结构方程系统。如果模型中有外生变量，还需对与它们的测量模型对应的矩阵（$\mathbf{\Lambda}_x$，$\mathbf{\theta}_\delta$）以及与结构效应对应的矩阵（$\mathbf{\Phi}$，$\mathbf{\Gamma}$）进行设定。

作为例子，我们将对二期德国抗议数据中，合法抗议倾向与群体资格之间的交叉滞后作用和同步因果作用进行检验，这一数据使用 3 个指标来测量抗议倾向。我们假定群体成员资格在第一轮、第二轮中无测量误差，由于两个时点都只有一个指标，这一变量可能的测量模型无法识别。因此，对 LISREL 的测量误差矩阵进行修正，以反映两个内生变量（$\text{GROUPS}_1$ 和 $\text{GROUPS}_2$）得到完美的测量，GROUP 和 PRPOTEST 潜变量的结构效应在 $\mathbf{B}$ 矩阵中设定，潜变量结构扰动项的方差和协方差在 $\mathbf{\Psi}$ 矩阵中设定。回忆前面的章节可知，如果没有其他仅影响被关注变量之一而不影响另一个的外部外生变量，同时带有交叉滞后作用和同步交互效应的二期模型就将不可识别。因此，要估计这一模型，必须为第二轮调查中的合法性抗议倾向和群体成员资格分别设置一个工具变量。这两个外生 $\xi$ 变量被设定只分别对两个交互相关的第二轮变量中的一个有影响。这样的变量并不是该研究数据收集的一部分，但调查含有受访者潜在的非法政治活动参与（如参与非法游行）的倾向以及受访者自报的、鼓励非法抗议行为的群体成员资格等信息。如果我们假定 1989 年的非法抗议倾向会影响 $\text{PROTEST}_2$ 但与 $\text{GROUP}_2$ 方程的扰动项无关，1989 年的非法成员资格会影响 $\text{GROUPS}_2$ 但与 $\text{PROTEST}_2$ 的扰动项无关，那么这些外部变量可被当做工具变量以识别这一模型。这个描述两轮调查间 GROUPS 和 PROTEST 关系的模型的最大似然估计为：

$$\text{PROTEST}_2 = 0.45\ \text{PROSTEST}_1 + 1.13\ \text{GROUPS}_2 + (-0.20)\ \text{GROUP}_1$$
$$\begin{matrix} (0.11) & (0.57) & (0.27) \\ \underline{0.41} & \underline{0.45} & \underline{-0.07} \end{matrix}$$

[4.14]

$$\text{GROUPS}_2 = 0.27\ \text{GROUPS}_1 + (-0.01)\text{PROTEST}_2 + 0.11\ \text{PROTEST}_1$$
$$\begin{matrix} (0.14) & (0.05) & (0.06) \\ \underline{0.25} & \underline{0.03} & \underline{0.25} \end{matrix}$$

这些估计结果表明,GROUPS 和 PROTEST 之间的关系与我们在利用三期全国青少年调查数据分析同龄伙伴涉入与违法行为之间的关系时,看到的反馈过程相同。抗议倾向对第二轮调查时的群体成员资格有显著的滞后作用(在控制群体成员资格初始值的情况下),而群体成员资格对抗议倾向具有同步作用。

### 利用背景变量来识别双指标测量模型

在一些情况下,并不存在 3 个指标用以估计带有相关的误差项的测量模型。在这种情况下,如果至少存在一个在两个时点都对潜变量有影响的背景变量,那么仍有可能识别图 4.2 中含有相关的测量误差的二期一双指标测量模型(Dwyer, 1983:400)。在 LISREL 中,可以通过将该背景变量作为外生 $\xi$ 变量纳入,并在 $\Gamma$ 矩阵中将其效应设定为自由参数的形式对这一模型进行设定。如果每轮调查中有一个 $\lambda$ 系数被设定为 1,那么现在全部自由参数的总数为 14:$\zeta_1$ 和 $\zeta_2$ 的方差、$\beta_{21}$、$\lambda_{21}$、$\lambda_{22}$、4 个误差方差及 2 个误差协方差、2 个从背景变量到潜变量的 $\gamma$ 作用,还有 1 个背景变量的 $\phi$ 方差。因为纳入背景变量后有 15 个 $(5 \times 6/2)$观察的方差和协方差,所以模型过度识别,带有 1 个自由度。

### 三期三指标模型

要识别和估计含有相关的测量误差的模型的最后一个办法,是至少再收集一轮信息。如果一个潜变量的两个指标被测量三轮,那么所有将潜变量与指标连起来的参数、潜变量的稳定性系数以及相邻轮次的指标的测量误差都能在没有限制

的情况下得到估计。在这种情况下，将有 21 个观察的方差协方差、18 个自由参数，因而整个模型带有 3 个自由度，可以被检验。该模型唯一的局限在于，如果没有进一步的限制，第一轮和第三轮指标的测量误差间的协方差将无法被识别。合理的限制之一是让相邻两轮调查的测量误差的协方差相等，在这一假定下，连接第一轮和第三轮指标测量误差的协方差也可被识别。另外，还可以将各轮调查的 $\theta_\varepsilon$ 值设定为相等，但与上面讨论过的单指标模型一样，这一限定欠缺合理性。

当潜变量的 3 个指标被测量的时点数达到 3 个甚至更多时，不相邻轮次的测量误差间存在相关的模型，在没有限定的情况下也能被估计。与调查轮次更少的数据相比，这一模型还允许对相关的测量误差进行更灵活的建模和解释。例如，Jagodzinski(1987)以及 Raffolovich 和 Bohrnstedt(1987)展示了带有相关的测量误差的模型是如何与经典的因子分析模型直接对应的，后者的每个观察变量都是下列 3 个方差的函数：公因子、纯误差方差以及具体题项因子——用来测量只属于该题项的方差，这一方差既非随机测量误差，也和公因子无关。因此，题项因子混合了前面讨论过的不同时点的测量误差间的相关。在截面分析中，具体题项方差与特有随机误差成分是混合在一起的，因而会高估一个指标的纯误差成分(Raffolovich & Bohrnstedt, 1987:386)。如果拥有多期面板数据，要想区分具体题项的方差的 3 个来源，就可以估计图 4.4 所示的模型。

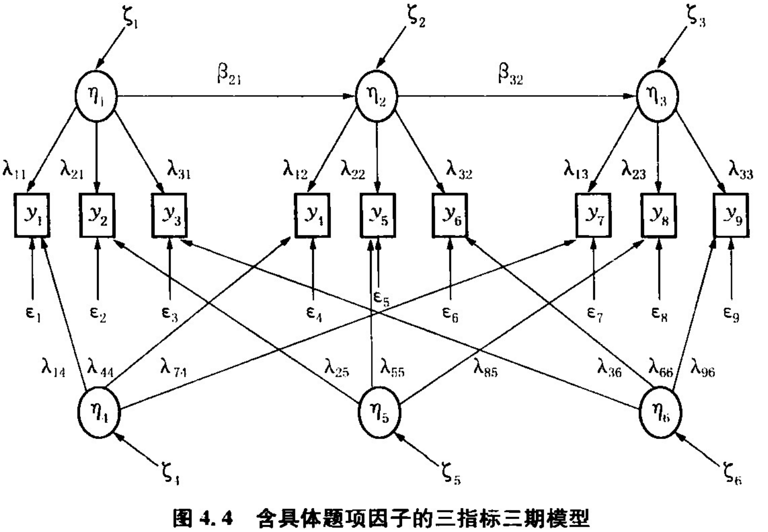



**图 4.4 含具体题项因子的三指标三期模型**

在这一模型中，$y_k$ 的每一个指标都由一个公因子（第一轮是潜变量 $\eta_1$，第二轮是 $\eta_2$，第三轮是 $\eta_3$）、一个误差项 $\varepsilon_k$ 以及一个具体题项因子（每轮调查的第一个题项对应的是 $\eta_4$，第二个对应的是 $\eta_5$，第三个对应的是 $\eta_6$）引起的。模型中有一个 λ 系数或因子负载将每个指标与实际关心的潜变量联系起来，另一个则在每个时点，将该指标连到题项因子上。这一模型假定，随机误差与结构扰动项与潜变量及具体因子无关，且假定具体因子与公因子彼此独立。

要设定前面在 LISREL 中估计过的简单的相关误差模型，只要让相邻期次的 $\varepsilon_k$ 间的协方差自由变化即可。由于这里包含具体因子，所以 LISREL 的设定在一定程度上更复杂了。各个公因子 $\eta$ 的刻度的设定是通过将每轮调查中的一个指标的系数固定为 1 来完成的，而具体因子的刻度同样是通过将某轮调查中连接具体因子和指标的系数固定为 1 来实现的。这些设定的选择是随意的，但每个潜变量一定要有一个指标的系数被固定为 1，各个具体因子也是如此。测量误差方差被设定互不相关，因为具体因子上的负载已经把握住了相同题项的历时相关。图 4.4 所展示的模型是过度识别的，它具有 16 个自由度，因此总共有 $9\times10/2$ 或者说 45 个观察的方差和协方差，而仅有 29 个自由参数需要估计。通过对某些系数施加限制，还可以对其他替代模型进行估计和比较。例如，不含具体题项因子的模型、公因子与其各指标之间的因子负载在不同时间相等的模型、具体因子与其各指标之间的因子负载在不同时期相同的模型都可以被检验和互相比较。只要这些模型相互嵌套，也就是说，模型所含变量相同而只是在 $\mathbf{\Lambda}$ 或 $\mathbf{\Theta}_\varepsilon$ 矩阵上的限制多出一个或几个，那就可以通过两个模型的 $\chi^2$ 之差对模型拟合的提高程度进行评估。

只要得到估计结果，就可能将每个指标的方差分解为共同部分、具体部分及纯随机误差部分。Raffolovich 和 Bohrnstedt（1987）表明，当所有潜变量（包括各具体因子）都被标准化成均值为 0、方差为 1 时，每个指标的总方差可以被估计为：

$$\hat{\sigma}^2(y) = \lambda_c^2 + \lambda_s^2 + \theta_\varepsilon \qquad [4.15]$$

其中，c 表示公因子，s 为具体因子。因此，要计算公因子、具体因子以及纯误差

方差各占题项方差总和的比例，就非常简单了。

最后，这些模型中的测量效应与本章讨论的所有模型一样，也可以放在被充分设定了的结构方程模型中进行估计。这一方法的优点是，潜变量稳定性的估计偏差将会降低，而且这些差异也可能影响测量误差的估计。另外，在很多分析中，校正测量误差是为了获得对潜变量之间因果作用的无偏估计，而同时对测量误差及结构效应进行设定，则能更好地实现这一目标。

# 第 5 章 | 虚假相关及自相关的扰动项

在前面的章节中,我们已经描述了通过估计带有各种时滞结构和变量测量误差的面板模型来强化因果推论的不同方法。不过,跟踪调查情境下的因果推论的成功不仅取决于准确设置的合适时滞结构以及对测量误差的矫正,还需控制那些对因果系统存在潜在污染效应的外部不可测量变量。不可测量变量是内生变量的结构扰动项出现自相关的主要原因,忽视这一因素将导致许多稳定得分面板模型中因果作用的 OLS 估计有偏。这个问题很普遍,但面板数据常常含有足够的信息对自相关进行控制,也能对变量间其他可能的虚假相关类型进行检验。虚假相关面板模型直接建立在前一章讨论过的测量模型基础上。

请认真回想一下我们使用“全国青少年调查”探讨过的违法同龄伙伴涉入程度(IDP)与自报违法行为(SRD)两个变量之间的关系。在第 3 章中,我们看到面板数据是如何有助于估计这两个变量间可能的交互作用的。不过,可能为真的另一个替代模型是,事实上 IDP 和 SRD 测量的并非完全不同的概念,相反,它们可能都是一个更一般的公因子的具体指标,如个体在不法行为方面的总行为倾向(Gottfredson & Hirschi, 1987)。在这种模型中,公因子“导致”受访者对两个变量的反应,而 IDP 和 SRD 量表之间的所有共变都能被它们与这一公因子的相关所解释。尽管依靠截面数据,我们无法排除公因子模型,但是在面板数据中,公因子模型是可以估计的,并且能与变量间存在直接因果作用的模型进行比较。

在另一些情况下,公因子模型能在理论基础上被排除。例如,我们有很好的理由假定对候选人的评价与个体的政党认同并非某个一般性概念的指标,关

于两个概念理论上的差异，已经有很好的论述了。但虚假相关的问题仍然存在，因为未被纳入模型的其他变量仍然可能对因果系统产生影响。这些变量并非政党认同和对候选人的评价所测量的概念的“潜在”变量，而是在数据收集时遗漏了的或者研究者根本未加以考虑的代表其他不同因素的变量。这些变量可能是背景性的社会或人口学的特征，例如之前服兵役的情况、之前的选举行为或者是对总统竞选表现的评价等态度因素，它们能够同时影响政党认同和对竞争候选人的评价。在截面研究中，这些变量都被归入内生变量的误差项，就这一点而言，不可测量变量与政党认同和候选人评分都有关，因此，连接这些变量的系数的估计将有偏。在跟踪调查中，这个问题会因不可测量变量存在一定的历时稳定性而更加复杂。在这种情况下，内生变量的结构扰动项会呈现自相关，模型参数的 OLS 估计也会不正确。在稳定得分面板模型中，不可测量变量会导致对滞后的 $Y$ 的作用估计有偏，$X$ 对 $Y$ 的作用的估计同样可能受其影响而出现偏差。

本章将介绍一些利用面板数据估计公因子和不可测量变量模型的可选方法。这些模型无法提供控制变量间虚假相关的万无一失的方法，它们都有着各自的前提假定，这些假定在一定的研究环境中可能成立，也可能不成立。尽管这些方法不会也不能产生实验设计所能提供的因果推论信心，但在提高非实验情境下的因果推论的质量上，它们确实提供了我们需要的帮助。

## 公因子模型

二期双变量公因子的一般模型如图 5.1 所示。在该模型中，$X$ 对 $Y$ 或 $Y$ 对 $X$ 都没有因果作用，观察到的两个变量间的协方差，是它们与公因子 $Z$ 在每个时点上的联合关系的函数。变量 $X$ 和 $Y$ 都还被一个假定与 $Z$ 不相关的随机测量误差项 $e_i$ 所影响。显而易见的是，该模型与第 4 章描述的双指标二期模型完全相同，图中的虚线表示 $X$ 和 $Y$ 的测量误差可能历时相关。该模型中需要估计的结构效应是包括公因子 $Z$ 在时点 1 和时点 2 上对 $X_1$ 和 $X_2$（$\lambda_1$ 和 $\lambda_2$）的作用、$Z$ 对 $Y_1$ 和 $Y_2$（$\lambda_3$ 和 $\lambda_4$）的作用以及 $Z$ 的稳定性（$\beta_1$）。为了便于表述，$X$、$Y$ 和 $Z$

被假定为标准化变量。对图 5.1 中的结构效应和测量误差作出不同假定而得到的各种虚假公因子模型，都可以用数据进行检验。

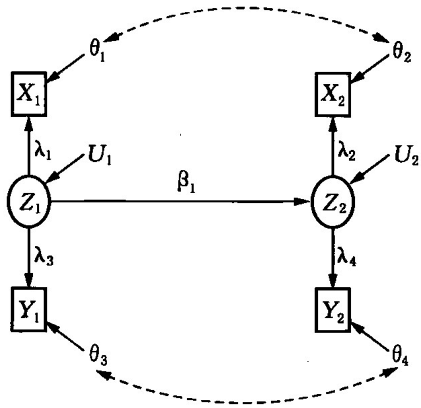



图 5.1 二期双变量公因子模型

两个带限制的公因子模型的检验可以直接通过比较具体观察相关系数的大小来进行。如果能够假定 $Z$ 对 $X_i$ 和 $Y_i$ 的因果作用在不同轮次中相等，也就是说，$\lambda_1=\lambda_2$ 且 $\lambda_3=\lambda_4$，并且 $X_1$ 与 $X_2$、$Y_1$ 与 $Y_2$ 之间不存在测量误差相关，那么模型预测结果中 $X_1$ 与 $Y_2$、$Y_1$ 与 $X_2$ 的交叉滞后相关将相等(Kenny, 1973)。正如第 3 章提到的，比较交叉滞后相关已不被认为是评估变量间因果优先次序的方法(Duncan, 1969; Rogosa, 1979; Shingles, 1976)，但它检验这种特定的(尽管是限制性的)虚假模型的作用仍然存在。当不存在相关的测量误差，且 $Z$ 对它的指标 $X$ 和 $Y$ 在不同调查轮次中的作用不相等时，对应的模型检验将变成：

$$r(X_1X_2)r(Y_1Y_2)=r(X_1Y_2)r(X_2Y_1) \qquad [5.1]$$

也就是说，取决于 $X$ 和 $Y$ 的历时稳定相关系数的乘积是否等于交叉滞后相关的乘积。

然而，这些检验常常是不完全的，因为它强行假定了调查轮次之间的测量误差相关。允许误差相关可能性的公因子模型可以在 LISREL 及相关框架里进行设定和检验，方法与前一章讨论的二期双变量测量模型一样。“全国青少

年调查"第一轮和第二轮的数据中，违法同龄伙伴涉入程度(IDP)与自报违法行为(SRD)之间的关系可以用来说明这些方法。

不含相关误差模型的 LISREL 初始估计显示，它无法解释 SRD 和 IDP 之间被观察到的历时相关。该模型强令 $Z_i$ 对 SRD 和 IDP 的结构效应在不同时期相等，从而导致对数据的拟合非常差，3 个自由度下的 $\chi^2$ 值为 161.3。放宽相等性限定后，模型大为改善，结果为 1 个自由度下的 $\chi^2$ 值等于 138.5。$\chi^2$ 值之差 22.8 在 2 个自由度下统计上非常显著，不过这一结果表明，模型的整体拟合度依旧很差。因此，在解释 NYS 数据中 SRD 和 IDP 的关系上，这两个限制模型都可能被拒绝。

不幸的是，含相关的测量误差的公因子模型是无法识别的，估计结构参数需要更多信息。一种办法在第 4 章中已讨论过，即增加一轮观察，所得的三期双变量模型将有足够的相关系数(15)用于结构参数的估计。然而，正如我们已经知道的，仅相邻轮次变量的测量误差相关的模型能在无限制的情况下得到识别。要估计第一轮和第三轮变量间的误差相关，更多指标或者对特定参数的相等性限定成为必要。不过，当只有两轮数据可用时，添加背景变量也能达到在无限制的情况下识别全部结构参数和测量误差相关的目的。该模型是过度识别的，带有 1 个可用以对模型的拟合度进行检验的自由度。

在当前的例子中，年龄作为观察的背景变量被纳入模型，可以重新对$IDP_1$和$IDP_2$之间以及$SRD_1$和$SRD_2$之间含有相关误差的模型进行估计。该模型比先前两种设定的模型的拟合度更好，1 个自由度下的 $\chi^2$ 是 4.9。尽管如此，这一拟合优度还不足以支持该模型在总体中成立。相比之下，依照先前发现的反馈过程(第二轮中的 IDP 对 SRD 具有同步作用，第一轮 SRD 对第 2 轮 IDP 具有交叉滞后作用)设定的 IDP 和 SRD 因果作用模型对数据的拟合极好，并且显示 IDP 和 SRD 之间的因果作用显著。在这个例子中，对虚假公因子模型的拒绝增强了我们对含有直接因果作用的模型适用性的信心。

当公因子模型未被拒绝时，它的估计系数的解释与上一章的测量模型中的完全一样。$Z_i$ 对 $X_i$ 和 $Y_i$ 的作用代表了这些指标对公因子的负载，$\beta_1$ 代表公因

子的历时稳定性。在加入了背景变量的模型中，$\gamma$ 系数代表背景变量对公因子的作用，而观察变量测量误差之间的相关代表 $X$ 或 $Y$ 的与公因子无关的具体题项的历时变化。

不过，和在前面几章中一样，需要强调的是，如果参数估计结果不合理或所表示的实际关系在理论上站不住脚，那么绝对不能仅仅根据统计结果而接受模型。在估计过程中应时刻注意模型背后的假定以及给定情境下，这些假定是否合理。如果模型并不拟合数据或在理论上不可接受，那么应该对其他替代模型，包括那些含有直接因果作用的模型，进行检验和比较。

## 不可测量变量模型

变量的虚假相关除了因为它们与同一个公因子有联合关系外，还可能是另一个过程造成的。$X$ 和 $Y$ 可能因为在既定模型中遗漏了理论上迥异的不可测量变量而出现相关。但如果关于这些不可测量变量的形式的某些假定能被证实，那么将不可测量变量的污染效应剔除后，面板数据就能用来估计所关注变量间的直接因果作用。

假定解释某个因变量 $Y_t$ 的真实模型为：

$$Y_t = \beta_0 + \beta_1 X_t + \beta_2 Y_{t-1} + Z_t + \varepsilon_t \qquad [5.2]$$

其中，$Z_t$ 是研究者未知的或不能被纳入模型的不可测量变量，$\varepsilon_t$ 是随机扰动项。不正确的估计方程变为：

$$Y_t = \beta_0 + \beta_1 X_t + \beta_2 Y_{t-1} + u_t \qquad [5.3]$$

其中，

$$u_t = Z_t + \varepsilon_t \qquad [5.4]$$

$Z_t$ 出现在扰动项 $u_t$ 中会产生一些重要的后果。首先，只要 $Z_t$ 的稳定性大于 0，$u_t$ 必然会历时相关。换句话说，外生变量 $Y$ 的结构扰动项会出现自相关。第二，如果 $Z$ 的稳定性不等于 0，方程 5.3 中的 $Y_{t-1}$ 将必然与 $u_t$ 相关，因为它们都与 $Z_{t-1}$ 相关。这意味着，OLS 不再适合用来估计这一模型，因为扰动项和解释

变量 $Y_{t-1}$ 之间的相关会导致方程参数估计的有偏和不一致（Johnston，1972），这通常会使 OLS 高估 $Y_{t-1}$ 对 $Y_t$ 的真实作用，因为被遗漏变量 $Z$ 的稳定性会导致 $Y$ 被观察到的稳定性中有一部分是虚假的。第三，一旦被遗漏变量 $Z_t$ 还与 $X_t$ 相关，那么新的偏差还会出现在对方程 5.3 的 $\beta_1$ 的估计中。在这种情况下，方程 5.3 中 $X_t$ 对 $Y_t$ 的作用至少是部分虚假的，因为它们都与 $Z_t$ 相关，偏差严重程度取决于 $Z$ 与这两个观测变量的相关强度。

与拥有大量历时性观察值的时间序列分析不同，在仅有少数几轮观察的短期面板研究中，校正因被遗漏变量而造成的偏差非常困难。当内生时滞变量作为预测变量时，这种问题最为严重，因为因果模型需要同时设定 $Y_{t-1}$ 对 $Y_t$ 和 $u_{t-1}$ 对 $u_t$ 的作用。显然，最好的方法是尽可能充分地设定实质性模型，这样一来，给定方程的误差项会相对较小，由被遗漏变量所引发的误差相关的危害也会降低。但设计局限往往会使这种策略不太可行，剩下的模型估计选择还有几种。

首先，可以将以外部观察变量形式出现的额外信息纳入模型，方程 5.3 的参数可能通过类似工具变量的估计方法得到复原。第二，可以对被遗漏变量 $Z$ 的稳定性作出不同假定，从而使结构扰动项之间的协方差以某种可预见的形式出现。通过在多轮面板分析中对扰动项的协方差施以限定，并在 LISREL 或类似的程序中估计模型，方程 5.3 的参数是有可能获取的。第三，可以在 LISREL 框架中对更多不可测量变量进行明确设定，并且如果信息——观察到的相关——充足，包含 $X$ 对 $Y$ 的结构效应以及 $Z$ 对每个变量的历时效应等在内的各种模型就都可以得到估计。我们将用本书已经分析过的经验案例来对这些方法逐个进行说明。

## 工具变量法

解决方程 5.3 中误差自相关问题的方法之一是使用工具变量提供额外信息，但它需要满足前几章已经提过的条件：(1) 与 $Y_{t-1}$ 相关，但是不与 $u_t$ 相关；(2) 对 $Y_t$ 不存在直接的因果作用。第一个前提条件保证工具变量与方程的扰动项不相关，第二个前提条件保证了模型可识别。如果这样的工具变量能够被找

到，那么二阶段最小二乘法或最大似然法都可以用来获得方程 5.3 的参数的一致估计。

例如，影响 1987 年和 1989 年的合法抗议倾向和群体资格的被遗漏变量就有可能导致前面章节所讨论模型的扰动项自相关。用来预测 1989 年的抗议行为和群体资格的稳定得分方程只有在两个滞后因变量——1987 年的合法抗议倾向和群体资格——存在一个或一组工具变量时，才能被估计出来。如果可以假定 1987 年的非法抗议倾向($Z_1$)影响了 1987 年的合法抗议倾向，但在控制其他变量的情况下，对 1989 年的合法抗议倾向没有直接的因果作用，且与 1989 年的合法抗议倾向的结构扰动项不相关，那么它就可以充当 1987 年合法抗议倾向的工具变量。TSLS 估计的程序与前面章节中的一样：用 $Y_{t-1}$ 对其工具变量和所有其他外生变量进行回归得到 $\hat{Y}_{t-1}$，然后用 $Y_t$ 对 $\hat{Y}_{t-1}$ 和在 $Y_t$ 预测方程中出现的所有其他变量进行回归。

如果交互效应模型中存在由被遗漏变量 $Z$ 引起的扰动项自相关，也就是说，方程 5.3 中的 $Y$ 也会影响 $X$，情况就会变得更复杂。在这种情况下，$X$ 和 $Y_{t-1}$ 都必然与 $u_t$ 相关，模型需要再纳入一个工具变量（Hannan & Young，1977）。在当前的例子中，如果我们假定时点 1 时，鼓励非法抗议的群体的资格($Z_2$)与时点 2 的结构扰动不相关，那它就可以被当做交互效应模型中的 GROUPS 的工具变量。

在 LISREL 框架中，这类模型的估计是前面估计过的模型的直接拓展。结构扰动项之间的有关协方差在 $\mathbf{\Psi}$ 矩阵中设定，如果 $PROTEST_1$、$PROTEST_2$、$GROUPS_1$、$GROUPS_2$ 是内生变量 1 到内生变量 4，就可以让参数 $\psi_{21}$ 和 $\psi_{43}$ 处于自由状态以允许自相关的可能性。工具变量被作为外生 $\xi$ 变量，它们的效应在含有外生变量和内生变量之间的结构效应的 $\mathbf{\Gamma}$ 矩阵中设定。在该矩阵中，为了保证模型的识别，必须对某些 $\gamma$ 效应加以限定。例如，将 1987 年的非法抗议倾向对 1989 年的 $PROTEST_2$ 的作用限定为 0，1987 年鼓励非法抗议的群体的资格对 1989 年 $GROUPS_2$ 的作用也设定为 0。其余效应和矩阵的设定与前面章节中的设定完全一样，具体设定取决于被假定掌控整个因果系统的准确结构和测量过程。

假定测量模型为前一章针对 $PROTEST_1$ 和 $PROTEST_2$ 的三指标测量模型，结构模型为在那里没有设置的“反馈模型”以揭示 GROUPS 和 PROTEST 之间历时性的结构效应，这一自相关模型是用 LISREL 最大似然法估计的，下面是模型的结构部分：

$$PROTEST_2 = \underset{\substack{(0.09)\\ \underline{0.29}}}{0.32}\ PROTEST_1 + \underset{\substack{(0.15)\\ \underline{0.46}}}{1.17}\ GROUPS_2 \qquad [5.5]$$

$$GROUPS_2 = \underset{\substack{(0.16)\\ \underline{0.86}}}{0.95}\ GROUPS_1 + \underset{\substack{(0.05)\\ \underline{-0.03}}}{(-0.01)}PROTEST_1$$

$PROTEST_1$ 和 $PROTEST_2$ 扰动项的协方差估计结果在统计上不显著，取值为 0.25，标准值是 0.05。而 $GROUPS_1$ 和 $GROUPS_2$ 两者的误差项的协方差估计值为−0.25，标准化取值是−0.34。这一结构扰动项之间的协方差估计在统计上显著。将本模型中的抗议倾向与群体资格之间的交互效应估计与方程 4.14 中的非自相关模型进行对比，就可以看到，此处两个变量间的关系完全是单向的，GROUPS 影响 PROTEST。两个模型中的 PROTEST 的稳定效应，尤其是 GROUPS 的稳定效应差异很大，这是不同时点的 GROUPS 变量的扰动项之间的协方差为较大负值的直接结果。

无论是通过二阶段最小二乘方法还是LISREL类方法来估计自相关扰动项的面板模型，都会出现这种类型的负向自相关（Long，1983b：78；Markus，1979:53）。尽管负向自相关存在具有实质性的解释的情况，但更可能意味着模型设定出错。一个很大的可能性是变量（在这里是 $GROUPS_1$ 和 $GROUPS_2$）测量无误差的假定是错误的。测量误差的存在和自相关，即自变量与其方程的误差项之间的相关，一样会给 OLS 估计带来相同的问题。由于测量误差的解决办法之一也是在分析中纳入工具变量，所以当这一偏差得到控制时，结构效应也常常会像在测量模型中的一样，不会变小。[16]因此，当负向自相关结果出现时，对模型的解释需要非常谨慎，并应该提出其他可以修正测量误差或设定错误的

新模型。

在只有二期观察时,工具变量法是控制自相关的扰动项的唯一可行办法,但有关工具变量的外生性和其对系统中其他变量的作用的潜在假定常常是站不住脚的。在当前的例子中,我们假定 1987 年的非法抗议倾向与 1987 年的合法抗议相关,但是对 1989 年的合法抗议没有影响,并且与 1989 年的合法抗议倾向的误差项中所有被遗漏的变量 $Z$ 都不相关,这一假定本身就可能是不合理的。而且,用工具变量替代滞后的内生变量通常会使所有因果作用的第二阶段估计变得很不准确。这些问题表明,工具变量方法作为控制因果系统中被遗漏变量的混淆效应的手段,在很多情况下是不太令人满意的。如果有三期或多期数据,下面介绍的方法通常更加适合。

## 对不可测量变量的稳定性作出各种假设

对带有自相关扰动项的因果作用进行估计的另一种方法是基于对方程 5.4 的误差项中被遗漏变量的稳定性的不同假设。在短期跟踪研究中,有理由假定 $Z$ 代表一些个体的或具体单位的特征,或者一些不随时间而改变的特性,也就是说,$Z_{t-1} = Z_t = Z$。因而方程 5.4 的误差结构变为:

$$u_t = Z + \varepsilon_t \qquad [5.6]$$

这里,$\varepsilon_t$ 是随机扰动项。在该假定之下,无论针对哪个间隔,方程 5.3 的误差项 $u_t$ 之间的相关都是定值,也就是说,时点 1 和时点 2 的误差项之间的相关与时点 1 和时点 3、时点 2 和时点 3、时点 1 和时点 4 等误差项之间的相关完全相同。在长期跟踪调查中,被遗漏变量更可能表现出不那么完美的历时稳定性,此时,误差项的结构将遵照时间序列研究中常见的一阶自回归模式:

$$u_t = \rho u_{t-1} + \varepsilon_t \qquad [5.7]$$

其中,$\rho$ 为取值小于 1 的自相关参数,$\varepsilon_t$ 和之前的定义一样。在这一误差结构下,误差项之间的相关会以几何级数衰减,如果相邻扰动项之间的相关为 $\rho$,那么相距两个间隔的扰动项之间的相关是 $\rho^2$,如此类推。

Tuma 和 Hannan(1984)以及 Stimson(1985)简单介绍了在混合横截面数

据和时间序列分析框架下，估计此类含有自相关误差结构的面板模型的方法。在三期或多期调查数据下，如果模型的因果结构不随时间而变化，并且相邻两次观察间的时间间隔相等，混合数据法将是合适的分析方法。在此处讨论的自相关模型中，这些方法的主要缺陷在于，和之前提到的一样，一旦两个变量被假定在一个交互因果模型中相互影响，就需要以工具变量的形式纳入更多信息。

尽管文献中的范例很少，但非混合的多期面板自相关模型也能在 LISREL 框架中进行估计。这种模型的估计与二期数据情况下一样，首先让结构扰动项的 $\Psi$ 矩阵的非对角元素自由取值，然后根据对扰动项结构的假定，给单个 $\psi$ 元素施加限制。在标准化的三期模型中，内生变量 $Y$ 有 3 个扰动项相关需要设为自由形式：$\psi_{21}$，代表第一轮、第二轮调查中 $Y$ 的扰动项的相关；$\psi_{32}$，代表第二轮、第三轮扰动项之间的相关；$\psi_{31}$，代表第一轮、第三轮扰动项的相关。类似的代表 $X$ 变量不同时间的误差项之间的相关参数也需要被放开。最后，带有恒定遗漏变量的模型可以通过使内生变量的扰动项之间的各个相关系数彼此相等的方式得到估计[17]，也就是说：

$$\psi_{21} = \psi_{32} = \psi_{31} \qquad [5.8]$$

表 5.1 展示了结构扰动项之间存在和不存在相关的两个模型的参数估计，使用的是上一章分析过的 1980 年的政党认同和候选人评价之间关系的例子。和往常一样，假定因果作用为交叉滞后设定，且限定第一轮到第二轮间 PID 对 THERM 的交叉滞后作用和第二轮到第三轮间的相等。THERM 对 PID 的作用以及 PID 和 THERM 的历时稳定性也都做了类似限定，对 PID 和 THERM 的误差相关做了如方程 5.8 那样的限定以模拟稳定的遗漏变量产生的自相关扰动项。

初始模型(1)设置了自相关扰动项，但是没有考虑任意变量的各个时点的测量误差。结果显示政党认同和民意评分估计的稳定性非常高，交叉滞后作用只在一个方向上——从 PARTY 到 THERM——统计显著。不过，和在两轮模型中一样，自相关估计值为负且在统计上显著，表明模型可能存在设定错误。

表 5.1　带有和不带自相关扰动项的三期模型

| | 模型(1) | 模型(2) |
|---|---|---|
| **稳定性** | | |
| 政党认同 | 1.0 | 0.99 |
| | (0.02) | (0.03) |
| 民意评分 | 0.80 | 0.71 |
| | (0.05) | (0.07) |
| **交叉一滞后作用** | | |
| 政党认同—民意评分 | 0.13 | 0.20 |
| | (0.03) | (0.04) |
| 民意评分—政党认同 | −0.02 | 0.01 |
| | (0.02) | (0.01) |
| **自相关的扰动项** | | |
| 政党认同 | −0.13 | −0.003 |
| | (0.01) | (0.03) |
| 民意评分 | −0.09 | 0.05 |
| | (0.03) | (0.04) |
| **同步的扰动项相关** | | |
| 政党认同—民意评分 | 0.04 | 0.03 |
| | (0.01) | (0.01) |
| **测量误差方差** | | |
| 政党认同 | | 0.13 |
| | | (0.03) |
| 　信度估计值 | | 0.87 |
| 民意评分 | | 0.20 |
| | | (0.04) |
| 　信度估计值 | | 0.80 |
| $\chi^2$(df) | 71.3(7) | 9.8(5) |
| | $p < 0.001$ | $p = 0.08$ |

注：表中数字为标准化的最大似然估计值。括号中为标准误。所有模型的 $N$ 都是 733。
资料来源：全国选举研究，主面板文件，1980 年。

考虑到上一章的结果模式，模型(1)无测量误差的假定很可能是错误的。模型(2)重新估计了该模型，但设置了第 4 章的三期单指标测量误差结构，并同时纳入自相关的扰动项。结果显示，两个自相关参数都很小，且在统计上不显著，而

稳定性和交叉滞后作用都与第 4 章发现的结果近似。此外，模型整体对数据的拟合要比无测量误差模型(1)好得多。我们对 1980 年大选期间，政党认同和候选人民意评分之间的关系为单向因果作用的信心更强了，因为在对因果系统中的自相关和测量误差假定不同的几个模型中，这两个变量间的作用模式一直保持相似。另外，在这个例子里，测量误差对因果作用估计的影响似乎比自相关的混淆效应大。尽管如此，自相关给因果推论带来的潜在错误仍然很严重，只要有可能，研究者就应当尽量使用不同的模型设定来验证确保因果结论的模式。

如果误差结构被假定服从方程 5.7 的一阶自回归模式，那么 LISREL 估计可以在标准化模型中运行。估计时，放开扰动项相关，并将相邻扰动项的协方差设定为彼此相等，然后令时点 1 和时点 3 的扰动项协方差等于时点 1、时点 2 的扰动项协方差与时点 2、时点 3 扰动项协方差的乘积。在 LISREL 中对模型参数进行此类限定的方法，请参见 Hayduk 的著作(Hayduk, 1987)。最后，还可以设置 LISREL 模型以对扰动项的协方差进行没有限定的估计，只要用来识别这些参数的信息足够。不过，在交互效应模型中，对这些相关进行无限定的估计通常是不可能的，而且这种特别的模型设定无法反映任何已知的误差项结构。

## 对不可测量变量明确建模

在 LISREL 或其他结构模型框架下估计带遗漏变量的模型的最后一种方法是明确设定一个额外的、没有测量指标的潜在变量，以代表因果模型中的不可测量变量 $Z$。为了使同时含有 $X$ 和 $Y$ 之间的结构效应以及来自“虚幻”变量 $Z$ 的效应的模型得到估计，至少需要 3 轮观察。这种方法的主要优点在于，它允许直接设定因果系统中测量变量和不可测量变量间假定的相互作用，并且只要有足够的自由度用以识别参数，它就能提供所有因果作用的大小和方向的最精细的估计。图 5.2 展示了这类不可测量变量模型中的一种可能形式，其中遗漏变量 $Z$(在 LISREL 中作为没有测量指标的新增 $\eta$ 来建模)一直同时影响 $X$ 和

$Y$，而 $X$ 和 $Y$ 之间也有交叉滞后作用。

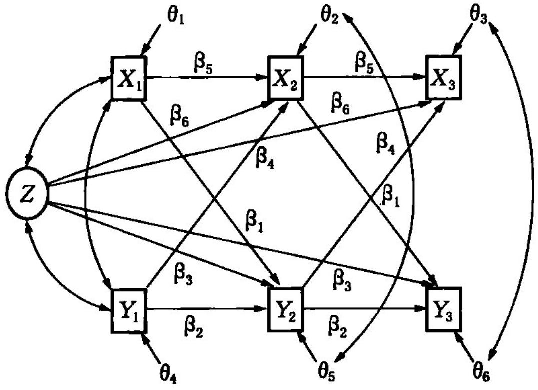



图 5.2 含有不可测量变量 $Z$ 的三期交叉滞后模型

**表 5.2 自报违法行为和违法同伴涉入度的三期不可测量变量模型**

| | |
|---|---|
| 结构效应 | |
| SRD 到 SRD | 0.40 |
| IDP 到 IDP | 0.14 |
| IDP 到 SRD | −0.04* |
| SRD 到 IDP | 0.01* |
| Z 到 SRD | 0.45 |
| Z 到 IDP | 0.58 |
| 时间 1 时的相关系数 | |
| r(SRD, IDP) | 0.54 |
| r(SRD, Z) | 0.61 |
| r(IDP, Z) | 0.63 |
| r(IDP, SRD 扰动项) | 0.13 |
| $\chi^2$(df) | 90.8(7) |
| $p$ 值 | <0.001 |

注：$N = 1725$。所有变量都已被标准化，表中所示的数字为最大似然估计。不带星号的系数都在 0.05 水平上显著。

资料来源：全国青少年调查（1976—1978 年）。

为了识别模型，必须限定结构效应在不同时期相等，$Y_2$ 和 $Y_3$ 方程以及 $X_2$ 和 $X_3$ 方程的误差方差也需分别设定彼此相等。不可观测变量 $Z$ 的方差被设为

1。与迄今估计过的其他交叉滞后模型一样,模型中纳入了时点 2 各变量扰动项之间的同步共变以及时点 3 各变量扰动项之间的同步共变,以便把握那些可能在同一轮而非下一轮调查中,对 $X$ 和 $Y$ 都造成影响的被遗漏因素。在此设定之下,模型拥有 7 个自由度,已经过度识别。随着观测期次的增加,需要限定的参数减少。同样值得注意的是,通过纳入那些与 $X$ 和 $Y$ 相互影响的外部观察变量,这个模型扩展起来非常容易。在实际操作中,这些变量将被当做图 5.2 的全部 6 个变量的外生作用来处理,使模型成为那些由测量和不可测量的外部变量造成的潜在虚假性的代表。为了便于说明,在此处不考虑这些变量。我们用前面分析过的青少年自报违法行为(SRD)—违法同伴涉入度(IDP)的关系,对这一模型进行了估计,其结果展示在表 5.2 中。

该模型对数据的整体拟合度很差,在 7 个自由度下的 $\chi^2$ 值是 90.8($p<0.001$)。但是回想一下第 3 章中含有 SRD 和 IDP 同步交互效应和交叉滞后作用的模型,在 8 个自由度下的 $\chi^2$ 值是 167.0,这表明,比起只含直接因果作用的模型,不可测量变量模型对数据的拟合度更好。有趣的是,模型显示,一旦控制了不可测量变量的影响,SRD 和 IDP 之间的交叉滞后作用在任意方向上都不再显著。重新估计一个带有 IDP 对 SRD 的同步作用以及 SRD 对 IDP 的滞后作用的不可测量模型,得到的 $\chi^2$ 值是 101.8,自由度是 8,与上面的交叉滞后模型的拟合度几乎相同。在这一"反馈"模型中,IDP 对 SRD 的标准化作用显著,取值为0.31,但 SRD 对 IDP 的标准化影响只有 0.10,比第 3 章所估计模型的值小。这两个模型在三期数据下无法直接比较,因为同时含有不可测量变量以及 $X$ 和 $Y$ 之间的同步因果作用和同步扰动相关的模型是不可识别的。

不管怎样,这些结论让我们难以继续认为两个变量间的交互作用的大小相似。数据与关系至少是压倒性的单向关系,或者说与其完全虚假这一解释相吻合。在此情况下,由于模型拟合度差,所以还应当估计并比较那些允许观察变量存在测量误差的其他可能设定。如果能通过新增观测轮次的方式提供更多信息,利用含有多个不可测量变量并允许不可测量变量存在历时变化的模型,我们就可以设定出能对这些解释进行更准确的相互检验的模型。

本章简要介绍了一些由公因子和不可测量变量造成的虚假相关模型和自相关模型。如果控制了所有这些可能的虚假性来源后,变量之间仍然存在显著的因果作用,那么研究者将会对结论的效度有更大的把握。不过应当强调,被观测变量造成的虚假相关也是跟踪调查设计中非常重要的问题,在得出因果结论之前,可以而且应该用数据对这些效应进行检验。显然,所有在理论上与 $X$ 和 $Y$ 有关的被观测变量都应被纳入模型。另外,我们可以估计那些剔除了所关注变量间的所有直接因果作用的模型,并从拟合度和实际合理性方面将其与因果作用模型加以比较。$X$ 和 $Y$ 之间的共变可能是由被观测的第三个变量或者由扰动项之间的同步共变产生,而不是由直接的因果作用产生。因为因果作用模型嵌套在没有因果作用的模型之中,它们的相对拟合度可以在 LISREL 或相关程序中通过 $\chi^2$ 差异得到估计。

所有这些模型都应被当做检验虚假相关的补充工具。应当强调的是,估计含有公因子和不可测量变量的模型不能完全替代研究者可获得的其他更确定的手段,即直接纳入那些被认为与给定因果系统有关的变量。不过,至少此处的模型试图将不可测量变量的作用明确地整合到模型设计中,而不是简单地假定(横截面数据中所做的就是这样)所有相关变量都已被纳入。此外,LISREL 方法在模型检验上具有很大的弹性,允许模型的相互检验以确定虚假相关可能出现的具体形式。在很多情况下,虚假相关可以通过这些检验被排除。如果没有被排除,来自直接因果作用模型的估计应该被谨慎对待,并在公因子或不可测量变量模型对数据的拟合同等有效这一事实基础上加以解释。

# 第 6 章 关于在面板分析中进行因果推论的结语

本书回顾了多种通过面板数据分析增强因果推论过程的方法。面板数据可以用来估计含有各种时滞设定、交互因果关系以及测量误差的模型，那些试图控制变量间潜在虚假相关的模型也可以得到估计。我们在整本书中已经表明，三期或更多轮次的面板数据可以使这些模型的估计和比较更方便，但在克服常见的成功因果推论的障碍上，相对截面数据，即便是二期面板数据也具有显著优势。

尽管比起截面和其他观测研究设计，面板数据分析有许多优点，但本书始终强调，面板数据不可能完全解决经验研究中因果推论方面的所有困难。此处开发的所有模型和方法都依赖于它们各自的假定，它们在特定情况下可能站不住脚或者会导致错误的经验结论。研究者应当时刻保持清醒，因为不同的估计模型可能都能拟合给定的数据，因而在特定研究背景中，没有哪个模型是唯一可接受的。正如 Dwyer(1983)提到的：

> 从被动的纵向数据中进行因果推论的智慧仅仅开始于人们意识到，一个被接受的模型与其他被同一套数据拒绝的多数模型同样有效。(1983:322)

另外，因果推论的过程并不只是设定并检验你想要证明的作用这么简单，在此过程中，研究者需要努力拒绝各种看似正确的模型，包括那些不含因果作用的模型和对直接因果假设进行检验的模型。这是一项极具挑战性的工作，对此，面板分析为非试验研究提供了当前可获得的、尽管不完美但却最有希望的工具。

# 附录

## LISREL 矩阵及符号系统

在 LISREL 结构模型中，内生变量用希腊字母 $\eta$ 指代，外生变量是 $\xi$，结构扰动项是 $\zeta$。将各个内生变量 $\eta$ 连接起来的结构效应被标记为 $\beta$ 系数，将外生变量(或先定变量) $\xi$ 与内生变量连起来的结构效应用 $\gamma$ 系数标记。例如，一个只含一个历时性地自我影响的变量的简单三期模型可以写成：

$$\begin{aligned} \eta_1 &= \gamma_{11}\xi_1 + \zeta_1 \\ \eta_2 &= \beta_{21}\eta_1 + \zeta_2 \end{aligned} \qquad [\text{A1}]$$

其中，$\xi_1$ 对应该变量第一轮调查的取值，$\eta_1$ 和 $\eta_2$ 分别为第二轮和第三轮的变量值。用矩阵形式表示，该系统如下所示：

$$\begin{pmatrix} \eta_1 \\ \eta_2 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ \beta_{21} & 0 \end{pmatrix} \begin{pmatrix} \eta_1 \\ \eta_2 \end{pmatrix} + \begin{pmatrix} \gamma_{11} \\ 0 \end{pmatrix} \xi_1 + \begin{pmatrix} \zeta_1 \\ \zeta_2 \end{pmatrix}$$

LISREL 结构方程系统的多方程矩阵一般形式是：

$$\boldsymbol{\eta} = \mathbf{B}\boldsymbol{\eta} + \boldsymbol{\Gamma}\boldsymbol{\xi} + \boldsymbol{\zeta} \qquad [\text{A2}]$$

这里，$\boldsymbol{\eta}$ 是 $m$ 个内生变量的向量，$\boldsymbol{\xi}$ 是 $n$ 个外生变量的向量，$\mathbf{B}$ 是连接内生变量的 $\beta$ 系数 $(m \times m)$ 矩阵，$\boldsymbol{\Gamma}$ 是一个 $(m \times n)$ 的连接外生变量和内生变量的 $\boldsymbol{\gamma}$ 系数矩阵，$\boldsymbol{\zeta}$ 是一个 $(m \times 1)$ 的结构扰动项向量。其他与估计目的相关的矩阵还有 $(n \times n)$ 的外生变量方差协方差矩阵 $\boldsymbol{\Phi}$ 以及 $(m \times m)$ 的结构扰动项的方差协方差矩阵 $\boldsymbol{\Psi}$。

当变量含有随机测量误差时，使用矩阵形式同时表示一个模型的测量方程和结构方程可使这个程序得到扩展。如果 $\eta$ 是内生变量的真实得分，$\lambda_y$ 是真实得分对观测指标 $Y$ 的回归系数，测量模型的多方程矩阵将会是：

$$\mathbf{y} = \mathbf{\Lambda}_y \boldsymbol{\eta} + \boldsymbol{\varepsilon} \qquad [A3]$$

上面的 $\mathbf{y}$ 是内生变量所有 $q$ 个指标的 $(q \times 1)$ 向量，$\mathbf{\Lambda}_y$ 是将 $m$ 个内生变量与对应指标连接起来的 $\lambda_y(q \times m)$ 矩阵，$\boldsymbol{\varepsilon}$ 是指标测量误差的 $(q \times 1)$ 向量。测量误差的方差协方差被汇拢在 $(q \times q)$ 的矩阵 $\boldsymbol{\theta}_\varepsilon$ 中，其单个元素是 $\theta_\varepsilon$。

当模型含有外生变量时，它们的每个指标可以被写成同样的一套测量方程。方程 A3 的对应方程为：

$$\mathbf{x} = \mathbf{\Lambda}_x \boldsymbol{\xi} + \boldsymbol{\delta} \qquad [A4]$$

其中，$\mathbf{x}$ 是外生变量 $\xi$ 的 $p$ 个指标的 $(p \times 1)$ 向量，$\mathbf{\Lambda}_x$ 是一个 $(p \times n)$ 的连接 $n$ 个外生变量与对应指标的 $\lambda_x$ 矩阵，$\boldsymbol{\delta}$ 是一个 $(p \times 1)$ 的指标测量误差向量。外生变量指标的测量误差的方差协方差都在 $(p \times p)$ 的矩阵 $\boldsymbol{\theta}_\delta$ 中，其单个元素为 $\theta_\delta$。

# 注释

[1] 面板数据的这些特点在其他形式的纵向数据中也能找到,"合并的时间序列数据"将多个单位的时间序列数据合并成单个的数据集(Sayrs, 1989)。尽管这两类数据没有形式上的区别,但合并的时间序列数据通常被认为含有更多时点,而面板数据含有更多的案例,它往往只包含大量案例的第二到第五轮数据。这种区别使对两种数据的分析策略也不尽相同(Kessler & Greenberg, 1981:第 11 章; Tuma & Hannan, 1984:第 2 章)。

[2] 见 Allison(1990)以及 Judd 和 Kenny(1981)关于在何种条件下,无条件的变化得分模型适合分析实验和准实验数据的讨论。

[3] Hargens、Raskin 和 Allison(1976:454—455)在一项关于科研产出的分析中认为,一个没有时滞 $Y$ 变量的模型"似乎特别适用于那些在每个时间间隔上都必须重新创造或生产的……变量,而不是那些具有内在稳定性原则的变量。换言之,除非对其采取行动,否则这些变量会倾向于保持不变的变量"。

[4] 其他含有滞后 $Y$ 变量作为自变量的理论模型还包括适应性预期模型和 Koyck 分布式滞后模型。在适应性预期模型中,因变量的真实值由自变量的未知"期望"值而非真实值决定,模型设定该期望值由上一时期的期望值的修正产生,修正量为最后一个时期的期望值与真实值之差的一个固定比例。而在 Koyck 模型中,$Y$ 的当前值由 $X$ 的当前值与所有过往值决定。这些模型通过代数运算,能产生一般形式与部分调整模型相同的估计方程,只是误差项结构不同,并引发了统计估计上的不同问题。关于这些模型及相关估计方法的具体介绍,见 Gujarati(1988)的著作。

[5] 正如 Plewis(1985:60) 所说:"($X_t$ 和 $Y_t$)之间的因果关联……与其说意味着同步作用,不如说意味着这样一种可能性,即 $X$ 在时点 1 之后某个时间发生的变化将导致时点 2 时 $Y$ 的变化,而所谓的'某个时间'没有任何精确的界定。"

[6] 由于 $X_1$、$X_2$、$\Delta X$ 中的每一个变量都是另外两个的线性组合,因而无法将 3 个变量同时纳入同一个回归模型以估计它们的独特效应。Kessler 和 Greenberg(1981:第 6 章)建议了一些估计全部效应的方法,但需要对三者的效应大小或它们的相互关系进行一些严格的假定,或者对该因果系统引入更多的信息。

[7] 这两个估计的一致性可被当做模型的整体检验之一。Tuma 和 Hanan(1984:344)指出,如果两个估计分歧很大,这表明不是初始微分方程的函数形式设定错误,就是外生变量 $X$ 的历时变化方式设定错误。

[8] SPSS 和其他计算机软件包中的 TSLS 程序提供了每个方程的结构扰动项估计,$U_1$、$U_2$ 之间的标准化协方差就是这些估计之间的相关系数。如果是通过 LISREL 的最大似然程序估计同步作用模型,那么模型带有的 1 个自由度允许对模型的假定进行正式的统计检验。如果 $\chi^2$ 值大于 3.84(即自由度为 1 时,0.05 显著性水平上的临界值),那么有关扰动项之间的共变为 0 的假定在总体中是不成立的。但是,不对这些参数进行约束,又会导致模型不可识别,除非纳入新的外生变量。

[9] Gillespie 和 Fox 的讨论严格限定在他们所谓的"平行"联立方程模型内,后者代表不同单位的相同变量之间的相互作用,例如某个青年的职业期望与其好朋友的职业期望是交互相关的。特别是在这些情况下,结构扰动项之间的负向共变在理论上是非常不可能的,但"内生变量和外生变量测量中的随机误差造成的后果和遗漏一系列外生变量的交叉—滞后作用的后果……(看起来)都适用于非递归模型"(Gillespie & Fox, 1980:304)。

[10] 通过一致性限定得到估计的三期模型的一个相关问题在于,这一模型要求因果系统在整个面板

期间尚未处于平衡状态。如果情况确实如此，那么更多期次的观察并不会提供更多有关变量协方差的新信息，因为第二轮至第三轮之间的协方差将等于第一轮、第二轮之间的取值。因而模型无法从经验上得到识别。Kessler 和 Greeberg(1981:37—46)表明，只有当系统远未达到平衡时，一致性限制这种方法才能产生精确的参数估计。

[11] 在标准化了的测量模型中，$\eta_t$ 的方差被固定为 1，所以 $\lambda_{it}$ 为一个自由参数。如果观察指标同样被标准化，那么 $\lambda_{it}$ 代表潜变量与观察指标之间的相关，这有时被称为"认知性相关"或"效度系数"(Sullivan & Feldman, 1979)。

[12] 如果随机误差出现在因变量中，对于自变量效应的 OLS 估计将会无偏但缺乏效率，因为方程的结构扰动项估计值将会被夸大。见 Berry 和 Feldman(1985:27—28)的著作。

[13] 如果 $y_t = \lambda_{it}\eta_t + \varepsilon_t$ 且所有变量都被标准化，那么，把这一方程平方，然后求其期望值，将得到 $1 = \lambda_{it}^2 + \delta_{\varepsilon_t}^2$，所以误差项的方差等于 $1 - \lambda_{it}^2$。

[14] 不过，Werts 等人(1971)指出，如果在 Wiley-Wiley 模型中不作出误差方差相等的假设，那么"中间"指标 $y_2$ 的误差方差可识别，其信度可以这样计算得到：$\rho_{y_2y_2} = r_{12}r_{23}/r_{13}$。但如果没有更多的限定，其他参数都将无法识别。

[15] 表中所示为潜变量和观察指标的方差都被标准化为 1 的"完全标准化解"。LISREL 也能产生一种仅仅只有潜变量的方差被标准化为 1 的标准化解。如果使用的是后一种结果，那么每个指标的信度应该这样计算：$\lambda^2/[\lambda^2 + \text{Var}(\varepsilon)]$。

[16] 在三期或多期模型中，其他可能的设定错误还包括：遗漏从第一轮调查的滞后内生变量到第三轮的内生变量的因果参数；遗漏从 $\varepsilon_1$ 到 $\varepsilon_3$ 的二阶自相关效应，还有其他对误差项结构的错误设定。所有这些模式还会使滞后的内生变量的稳定效应往高的方向偏，这在未设定这些效应的模型中，体现为负向自相关。如果模型中有足够的自由度，那么这些效应都可以被纳入 LISREL 并进行检验。

[17] 这种类型的三期模型可能存在一些偏差，因为时间 1 的扰动项含有时间 1 变量的所有变异，而时间 2 和时间 3 的扰动项仅含有残余的变异。如果有 4 轮或更多轮次的调查数据，扰动项的协方差或相关相等的假定可以设定在时间 2、时间 3、时间 4 等的扰动项之间。

# 参考文献


Allison, P. D. (1990). "Change scores as dependent variables in regression analysis." In C. C. Clogg (ed.), *Sociological Methodology 1990*. Oxford: Basil Blackwell.

Arminger, G. (1986). "Linear stochastic differential quation models for panel data with unobserved variables." In N. B. Tuma (ed.), *Sociological Methodology 1986*. San Francisco: Jossey-Bass.

Arminger, G. (1987). "Misspecification, asymptotic stability and ordinal variables in the analysis of panel data." *Sociological Methods & Research 15*: 336—348.

Asher, H. B. (1983). *Causal Modeling*. Sage University Paper series on Quantitative Applications in the Social Sciences, 07—003. Beverly Hills, CA: Sage.

Bentler, P. M. (1985). *Theory and Implementation of EQS: A Structural Equations Program*. Los Angeles: BMDP Statistical Software, Inc.

Berry, W. D. (1984). *Nonrecursive Causal Models*. Sage University Paper series on Quantitative Applications in the Social Sciences, 07—037. Beverly Hills, CA: Sage.

Berry, W. D. & Feldman, S. (1985). *Multiple Regression in Practice*. Sage University Paper series on Quantitative Applications in the Social Sciences, 07—050. Beverly Hills, CA: Sage.

Bohrnstedt, G. W. (1969). "Observations on the measurement of change." In E. F. Borgatta (ed.), *Sociological Methodology 1969*. San Francisco: Jossey-Bass.

Bollen, K. (1989). *Structural Equations With Latent Variables*. New York: Wiley.

Carmines, E. G. & Zeller, R. (1979). *Reliability and Validity Assessment*. Sage University Paper series on Quantitative Applications in the Social Sciences, 07—017. Beverly Hills. CA. Sage.

Coleman, J. S. (1968). "The mathematical study of change." In H. Blalock & A. Blalock (eds.), *Methodology in Social Research*. New York: McGraw-Hill.

Costner, H. L. (1969). "Theory, deduction and the rules of correspondence." *American Journal of Sociology 75*: 245—263.

Duncan, O. D. (1969). "Some linear models for two wave, two variable panel analysis." *Psychological Bulletin 72*: 177—82.

Dwyer, J. H. (1983). *Statistical Models for the Social and Behavioral Sciences*. New York: Oxford University Press.

Elliott, D. S., Huizinga, D. H. & Ageton, S. S. (1985). *Explaining Delinquency and Drug Use*. Beverly Hills. CA: Sage.

Feldman, S. (1989). "The reliability and stability of policy positions: Evidence from a five-wave panel study." In J. A. Stimson (ed.). *Political Analysis*. Ann Arbor: University of Michigan Press.

Finkel, S. E. (1993). "Reexamining the 'minimal effects' model in recent presidential elections." *Journal of Politics 55*: 1—21.

Finkel, S. E., Muller, E. N. & OPP, K.-D. (1989). "Personal influence, collective rationality, and mass political action." *American Political Science Review 83*: 885—903.

Gillespie, M. W. & Fox, J. (1980). "Specification error and negatively correlated disturbances in 'parallel' simultaneous-equation models." *Sociological Methods & Research 8*: 273—308.

Gottfredson, M. & Hirschi, T. (1987). "The methodological adequacy of longitudinal research on crime and delinquency." *Criminology 25*: 581—614.

Green, D. & Palmquist, B. (1990). "Of artifacts and partisan stability." *American Journal of Political Science* 34: 872—902.

Gujarati, D. (1988). *Basic Econometrics*. New York: McGraw-Hill.

Hannan, M. T. & Young, A. A. (1977). "Estimation in panel models: Results on pooling cross-sections and time series." In D. R. Heise(ed.), *Sociological Methodology 1977*. San Francisco: Jossey-Bass.

Hargens, L. L., Raskin, B. F. & Allison, P. D. (1976). "Problems in estimating measurement error from panel data: An example involving the measurement of scientific productivity." *Sociological Methods & Research 4*: 439—458.

Hayduk, L. A. (1987). *Structural Equation Modeling With LISREL: Essentials and Advances*. Baltimore: Johns Hopkins University Press.

Heise, D. R. (1969). "Separating reliability and stability in test-retest correlation." *American Sociological Review 34*: 93—101.

Hendrickson, L. & Jones, B. (1987). "A study on longitudinal causal models comparing gain score analysis with structural equation approaches." In P. Cuttance & R. Ecob(eds.), *Structural Modeling by Example*. Cambridge: Cambridge University Press.

Jagodzinski, W., Kühnel, S., M. & Schmidt, P. (1987). "Is there a 'Socratic effect' in nonexperimental panel studies?" *Sociological Methods & Research 15*: 259—302.

Jennings, M. K. & Niemi, R. (1975). "Continuity and change in political orientations: A longitudinal study of two generation." *American Political Science Review 69*: 1316—1355.

Johnston, J. (1972). *Econometric Methods*. New York: McGraw-Hill.

Jöreskog, K. G. & Sörbom, D. (1976). "Statistical models and methods for analysis of longitudinal data." In D. J. Aigner & A. S. Goldberger(eds.), *Latent Variables in Socioeconomic Models*. Amsterdam: North Holland.

Jöreskog, K. G. & Sörbom, D. (1989). *LISREL 7: A guide to the Program and Applications*. Chicago: SPSS.

Judd, C. M. & Kenny, D. A. (1981). *Estimating the Effects of Social Interventions*. Cambridge: Cambridge University Press.

Judd, C. M., Krosnick, J. & Milburn, M. A. (1981). "Political involvement and attitude structure in the general public." *American Sociological Review 46*: 660—669.

Kenny, D. A. (1973). "Cross-lagged and synchronous common factors in panel data." In A. S. Goldberger & O. D. Duncan(eds.), *Structural Equation Models in the Social Sciences*. New York: Seminar Press.

Kessler, R. C. & Greenberg, D. F. (1981). *Linear Panel Analysis*. New York: Academic Press.

Liker, J. K., Augustyniak, S. & Duncan, G. J. (1985). "Panel data and models of change: A comparison of first difference and conventional two-wave models." *Social Science Research 14*:

80—101.

Long, J. S. (1983a). *Confirmatory Factor Analysis*. Sage University Paper series on Quantitative Applications in the Social Sciences, 07—033. Beverly Hills, CA: Sage.

Long, J. S. (1983b). *Covariance Structure Models: An Introduction to LISREL*. Sage University Paper series on Quantitative Applications in the Social Sciences, 07—034. Beverly Hills. CA: Sage.

Markus, G. B. (1979). *Analyzing Panel Data*. Sage University Paper series on Quantitative Applications in the Social Sciences, 07—018. Beverly Hills, CA: Sage.

Markus, G. B. (1982). "Political attitudes in an election year." *American Political Science Review 76*: 538—560.

Mayer, L. S. & Carroll, S. S. (1987). "Testing for lagged, cotemporal and total dependence in cross-lagged panel analysis." *Sociological Methods & Research* 16: 187—217.

McAdams, J. (1986). "Alternatives for dealing with errors in the variables: An example using panel data." *American Journal of Political Science* 30: 256—278.

McArdle, J. J. & Aber, M. S. (1990). "Patterns of change within latent variable structural equation models." In A. Von Eye(ed.), *Statistical Methods in Longitudinal Research. Volume 1: Principles and Structuring Change*. San Diego: Academic Press.

McDonald, R. P. (1980). "A simple comprehensive model for the analysis of covariance structures: Some remarks on application." *British Journal of Mathematical and Statistical Psychology 33*: 161—183.

Menard, S. (1991). *Longitudinal Research*. Sage University Paper series on Quantitative Applications in the Social Sciences, 07—076. Newbury Park. CA: Sage.

Menard, S. & Elliott, D. S. (1990). "Longitudinal and cross-sectional data collection and analysis in the study of crime and delinquency." *Justice Quarterly 7*: 11—55.

Nesselroade, J. R., Stigler, S. M. & Baltes, P. B. (1980). "Regression toward the mean and the study of change." *Psychological Bulletin 88*: 622—637.

Ostrom, C. W. (1978). *Time Series Analysis: Regression Techniques*. Sage University Paper series on Quantitative Applications in the Social Sciences, 07—009. Beverly Hills, CA: Sage.

Palmquist, B. & Green, D. (1989, September). "Estimation of correlated measurement errors in panel data." Paper presented at the annual meeting of the American Political Science Association, Washington. DC.

Plewis, I. (1985). *Analysing Change: Measurement and Exploration Using Longitudinal Data*. Chichester. UK: John Wiley.

Porst, R., Schmidt, P. & Zeifang, K. (1987). "Comparison of subgroups by models with multiple indicator." *Sociological Methods & Research 15*: 303—315.

Raffolovich, L. E. & Bohrnstedt, G. W. (1987). "Common, specific and error variance components of factor models: Estimation with longitudinal data." *Sociological Methods & Research 15*: 385—405.

Rindskopf, D. M. (1984). "Using phantom and imaginary latent variables to parametrize constraints in linear structural models." *Psychometrika 49*: 37—47.

Rogosa, D. (1979). "Causal models in longitudinal research: Rationale, formulation. and interpretation." In J. R. Nesselroade & P. B. Baltes (eds.), *Longitudinal Research in the Study of Behavior and Development*. New York: Academic Press.

Sayrs, L. W. (1989). *Pooled Time Series Analysis*. Sage University Paper series on Quantitative Applications in the Social Sciences, 07—070. Newbury Park. CA: Sage.

Shingles, R. D. (1976). "Causal inference in cross-lagged panel analysis." *Political Methodology 3*: 95—133.

Stimson, J. A. (1985). "Regression in space and time: A statistical essay." *American Journal of Political Science 29*: 914—947.

Sullivan, J. & Feldman, S. (1979). *Multiple Indicators: An Introduction*. Sage University Paper series on Quantitative Applications in the Social Sciences, 07—015. Beverly Hills. CA. Sage.

Tuma, N. & Hannan, M. (1984). *Social Dynamics*. New York: Academic Press.

Werts, C. E., Jöreskog, K. G. & Linn, R. L. (1971). "Comment on 'The estimation of measurement error in panel data'." *American Sociological Review 36*: 110—113.

Wheaton, B., Muthén, B., Alwin, D. F. & Summers, G. F. (1977). "Assessing reliability and stability in panel models." In D. R. Heise(ed.), *Sociological Methodology 1977*. San Francisco: Jossey-Bass.

Wiley, D. E. & Wiley, J. A. (1970). "The estimation of measurement error in panel data." *American Sociological Review 35*: 112—117.

Wiley, J. A. & Wiley, M. G. (1974). "A note on correlated errors in repeated measurements." *Sociological Methods & Research 3*: 172—188.

# 中介作用分析

## 作者简介

### 道恩·亚科布齐（Dawn Iacobucci）

美国范德比尔特大学管理学系欧文研究生院市场营销学教授。她的研究集中在网络、顾客满意度和统计方法上，并在各种市场营销学和心理学期刊上发表过多篇论文，还是《市场营销管理》（*Marketing Management*）一书的作者，并且与吉尔伯特·丘吉尔（Gilbert Churchill）合著了《市场学研究》（*Marketing Research*）。

## 译者简介

### 李骏

2009年获香港科技大学社会科学部硕士学位，目前是香港科技大学社会科学部博士候选人。研究兴趣包括中国城市社会分层与劳动力市场。

# 第 1 章 | 中介作用简介

中介作用分析是用来考察某项数据是否具有中介结构的一套统计方法。中介结构这一概念指向这样一种特定形式的因果关系机制——某个自变量对某个因变量的影响可能并非是直接的,而是通过某种中介过程实现的,或者说,某个自变量是通过某个中介变量而间接影响某个因变量的。试图检验中介关系的研究者通常将这些变量之间的相互关系称为"因果关系",例如,自变量影响了中介变量(或自变量的变化导致了中介变量的变化),而中介变量又影响了因变量。研究者随后会用一套统计方法来检验这些变量在经验数据中的相互关系。在本书中,我们将同时从概念和统计两个方面来讨论中介关系的检验方法。

在本章,我们将首先回顾确立和证实因果关系的逻辑以及(在社会学这一学科内人们所认同的)方法。不管使用哪种具体的统计工具来检验中介模型,这些概念层次的讨论都是切题的。

在中介关系研究文献中,一些学者已经重新强调了要注意对数据结构中的相关关系所作出的因果关系解释(例如,Holland, 1986; James & Brett, 1984; James, Mulaik & Brett, 1982; McDonald, 2002),其中更有一些人认为,实验方法仍然是最理想的确立因果关系的研究设计(例如,Shrout & Bolger, 2002; Spencer, Zanna & Fong, 2005)。一些研究者已经尝试对基本的中介统计方法进行改进或提升(例如,Kenny, Kashy & Bolger, 1998; MacKinnon, Lockwood, Hoffman, West & Sheets, 2002; MacKinnon, Warsi & Dwyer, 1995),还有一些研究者已经同时讨论了因果逻辑层次的议题和经验层次的方法改进问题(例如,Bentler, 2001; Cote, 2001; Lehmann, 2001; McDonald, 2001; Netemeyer, 2001)。我们将讨论所有这些主题。

## 中介作用是什么以及它为什么重要?

许多社会科学研究者都对中介作用感兴趣(Baron & Kenny, 1986; Iacobucci, 2001; James & Brett, 1984; MacKinnon et al., 2002)。中介作用背后的理论假设是,中介变量反映了自变量通过它来影响因变量的过程。研究者试图评估自变量对因变量的直接作用或(通过中介变量产生的)间接作用有多大。

如图 1.1 所示,$X$ 是自变量,$M$ 是假定的中介变量,$Y$ 是因变量。传统上来说,研究者是要考察预测变量(例如 $X$)与因变量(例如 $Y$)之间的关系。但在假定和检验中介作用的研究中,研究问题则变成了 $X$ 对 $Y$ 的作用是直接的(即 $X \rightarrow Y$)还是间接的(通过 $M$ 起作用,即 $X \rightarrow M \rightarrow Y$)? 间接作用情况下的假设是,变量 $X$ 有助于预测和解释中介变量 $M$ 的变化,而这又有助于预测和解释变量 $Y$ 的变化。

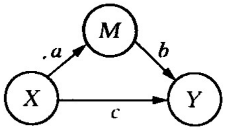



**图 1.1 简单的、标准的三变量中介模型**

探寻中介作用的目的是理解 $X$ 影响 $Y$ 的具体过程。在发现了某些经验现象(即 $X$ 看似影响 $Y$)之后,通过理解这一经验发现的本质(即 $X$ 如何影响 $Y$),才能实现理论的发展和提升。这种解释或许很简单:$X$ 本身确实部分地解释了 $Y$ 的变化。但或许还存在另外一种替代性的解释:$X$ 之所以能部分地解释 $Y$ 的变化,只不过是因为 $X$ 和 $Y$ 都与某个中介变量 $M$ 相关。

研究者似乎发现,中介作用逻辑从直观上来看,相当具有吸引力。反映中介作用重要性的一个度量是有关中介作用的检验和逻辑的学术论文出现的数量。确实,检验中介作用的文献非常之多,下面只是列举了一些最近的应用实例:

第一,$X$ 可以是某种特性(例如,认知需要);$M$ 是态度(例如,对广告的态度);$Y$ 是判断或应对(例如,购买某种广告商品的可能性)。

第二，$X$ 可以是某种情绪诱导；$M$ 是认知评估；$Y$ 是接受外界刺激后的感知测验结果。

第三，$X$ 可以是投票者的年龄；$M$ 是投票者的政治倾向（自由主义或保守主义）；$Y$ 是投票倾向。

第四，$X$ 可以是国家的经济发展条件；$M$ 是国家民主治理程度的测量指标；$Y$ 是国家的繁荣程度。

第五，$X$ 可以是对个人冒险倾向的测量；$M$ 是对某种特定金融产品的风险评级；$Y$ 是个人投资这种金融产品的倾向。

第六，$X$ 可以是对文化的测量（例如，个人主义）；$M$ 是个人风格与公司文化之间的适应性指标；$Y$ 表示领导才能。

第七，$X$ 可以是家庭收入；$M$ 是对学校环境的测量（例如，促进自主学习）；$Y$ 是学生在标准测试中的成绩表现。

上述主题各异的研究都关注中介作用，这支持了前文的观点，即中介作用相当重要，因为对它的研究很流行。姑且不论研究的内容和理论立足点是什么，对研究者来说，检验中介作用之所以具有吸引力，原因在于它试图追踪 $X$ 影响 $Y$ 的具体过程。

我们在后文即将讨论检验中介作用是否存在的有关方法论问题，例如，中介作用分析在当前的研究文献中是如何进行的？这些分析如何做得更好？在棘手情况下如何处理？但是，让我们先来回顾一下涉及因果关系推论的几个概念性的问题。

## 因果关系

本节我们将回顾表明因果关系确实存在的几个必要条件。这样做的原因在于，检验中介作用的研究者经常用因果关系的语句来表达他们的发现，例如说："$X$ 通过 $M$ 影响 $Y$。"或者有时更明确地说："$X$ 影响 $M$，而 $M$ 又影响 $Y$。"在解释如图 1.1 那样的路径模型时，对从 $X$ 到 $M$ 到 $Y$ 的因果关系链不加任何说明或推论，几乎已经成为了一种不可抗拒的习惯。在本节的讨论中，我们将承

认依据中介分析方法来确立因果关系所面临的挑战。正如下文所述，这涉及两个问题。首先，用来检验和分析中介作用的数据通常只是相关性的，而用它来得出有关因果关系的清晰结论在数据上具有局限性。确实，许多社会科学家都认为，实验方法中对变化机制的操控以及测量方法的使用仍然是最好的，并且，或许只有实验方法才能真正洞察因果关系。其次，任何一种统计方法本身都无法从相关性数据中“提取”它所潜藏的因果关系含义。然而，表明因果关系仍然是许多检验中介作用过程的研究者的目标，因此，一些指导性的原则对于理解各种情形下因果关系何以成立以及提高因果关系成立的可能性来说相当重要。

在试图用证据来确立因果关系时，科学哲学家都趋于接受 Hume 和 Mills 所提出的必要性准则：(1)共变性；(2)次序性；(3)排除竞争性的解释。我们将对此分别进行讨论。

### 共变性

共变性准则是指，如果 $X$ 是 $Y$ 的原因，则 $X$ 与 $Y$ 必然相关。众所周知，社会科学家的研究对象——人类的行为方式，十分微妙和复杂，因此，即使满足这第一个要求也是有困难的。

这牵涉到几个问题。首先，可能存在多种原因，也就是说，$X$ 可能只是 $Y$ 的几个原因中的一个(例如，家庭收入虽然对汽车消费有影响，但它并不是汽车消费的唯一决定因素)。在十分复杂的情况下，$X$ 与 $Y$ 之间的相关性可能会很低，因为还有其他因素可以解释 $Y$。

第二，即使关于 $X$ 与 $Y$ 之间关系的理论假设在研究总体中确实成立，众所周知，用抽样以及特定的测量方法或变量来对 $X$ 和 $Y$ 进行操作时，也都会不可避免地造成随机误差。由于误差的存在，在经验数据中所观察到的表示 $X$ 和 $Y$ 的变量之间的关系会被减弱或冲淡。

第三，如果两个变量之间的相关性出于上述原因或其他任何原因而被削弱，由于中介作用分析所关心的间接作用是两个概率函数——分别涉及 $r_{XM}$ 与 $r_{MY}$——的联合作用结果，那么在经验数据中发现可以用来支持间接作用路径的

结果的可能性就会变小，相应地，失败的可能性就会增大。也就是说，如果全部的作用路径在估计时都存在误差，那么唯一的那个直接作用路径会显得更为突出，因为间接作用路径涉及的两对关系所存在的双重误差对结果更具干扰性。

第四，对社会科学家来说，似乎永远都要小心避免牵强地将相关数据解释为因果关系。共变性的逻辑是，如果存在因果关系，那么就应该存在相关关系。即便是学习基本统计的学生都知道，与这句话相反的表述并不一定正确，也就是说，相关关系并不一定意味着因果关系。换言之，$X$ 与 $Y$ 之间的相关可能是因为 $X \rightarrow Y$，也可能是因为 $Y \rightarrow X$，还可能是因为有第三个变量同时影响了 $X$ 和 $Y$。因此，相关性的存在并不能使我们对这些情况中的任何一种作出定论。这种提醒并非多余之举，因为使用相关性数据来论证因果关系的情况在研究者分析中介作用的发表论文中（甚至在顶级期刊中）相当普遍。Baumrind（1983）在关于相关性数据不能作为因果关系结论的依据这一点上，立场十分鲜明："我们应该打消统计方法能够证明从相关性数据中可以得出因果关系推论的正当性的念头。"（1983：1289）Breckler（1990）有类似的言论："相关性数据的确不能为因果关系推论提供充足的依据。"（1990：269）

实验方法在微观社会科学（与社会学或宏观经济学相对，例如，消费行为学、心理学）中甚为流行，它被普遍认为是识别因果关系的最精确的方法设计（Mill 主张用实验方法来确立关键因素之间的因果关系，并要在实验设计中保证观察个体的随机分组；另见 Holland，1988；Rubin，1974、2005）。"传统实验方法的目标是要保证实验者能够清楚地推论因果关系是否存在以及它的作用方向。"（Baumrind，1983：1290）既然实验方法能够更好地确立因果关系，Spencer 等人（2005）就提出，对因果关系链（$X \rightarrow M \rightarrow Y$）的检验应该通过一系列实验方法进行。具体来说，在一个实验设计中，研究者通过操控自变量 $X$ 的变化来测量它对 $M$ 的影响；在另外一个实验设计中，则通过操控 $M$ 的变化来测量它对 $Y$ 的影响。[1]

中介作用分析有时确实会在实验环境中进行：先做一个实验，再通过方差分析（ANOVA）来看因变量的变化。即使研究者只用一个简单的 2×2 实验设计，并用 ANOVA 来分析所得结果，对某一因果关系假设的检验就已达到了实

验方法所能够达到的效果。[2]如果这一简单的 2×2 实验设计已经达到了目的，就确立因果关系而言，实验者应该就此停止，因为它已经达到了作为一种检验方法所需的严谨性。加入中介作用分析并不会使分析方法的严谨性更高，也不能证明研究者有多大的信心可以就因果关系下结论。但实际情况是，有时候研究者会在对关键因变量进行 ANOVA 分析后，再将一些补充性的测量加入中介作用分析。可以想象，研究者相信通过引入其他的统计检验（即与中介作用分析有关的一些统计指标）可以增加其分析的价值或严谨性，但事实并非如此，至少就确立因果关系而言，并非如此。当研究者加入中介作用分析时，他们其实是在分析相关性数据，而这种数据在识别因果关系时，却并不具备任何优越性。因此，在报告对核心因变量所进行的 AVONA 分析结果之后再加入中介作用分析，对一篇论文来说其实是减分而不是加分。所以，中介作用分析不应该是研究者的一个下意识的动作。

有了实验数据，研究者就有了较强的逻辑基础来论证 $X \to M \to Y$ 这一因果关系（例如，Netemeyer，2001）。对这种具有高度内部效度的研究来说，除了研究者刻意操控的因素之外，所有其他的条件都不变，这样一来，因变量的变化就可以被纯粹地归因于实验干预项的作用。然而，如果是对横截面数据或相关性数据进行中介作用分析，研究者的措辞就必须有所保留，或许可以说是“结构性的关系”，但不能说是“因果性的关系”。

## 次序性

次序性的准则是指，如果 $X$ 是 $Y$ 的原因，则 $X$ 一定是在 $Y$ 之前发生的。要满足这一条件似乎很简单。在理想状态下，研究者应该按变量在理论所预示的过程或机制中出现的顺序来依次测量：首先是 $X$，其次是 $M$，最后是 $Y$。这里所说的依次测量不是简单地指它们出现在同一个调查问卷中的不同页面上，而是指它们出现在 3 次数据收集进行时所处的 3 个时间点上（$t_1$、$t_2$、$t_3$）。这种测量顺序的存在使考察因果链或时间链的动态过程成为可能。[3]在时间点 $t_1$ 测量的变量有可能影响在时间点 $t_2$ 测量的变量，但在时间点 $t_2$ 测量的变量却绝不可能影响在时间点 $t_1$ 测量的变量（实验方法也是依时间维度进行的，对某个变量

或条件的操控发生在时间点 $t_1$，对因变量的测量发生在时间点 $t_2$，即使两个时间点之间的间隔时间很短）。

次序性是可以实现的，最理想的办法就是让受访者在时间点 $t_1$ 回答有关 $X$ 的调查题目，在时间点 $t_2$ 回答有关 $M$ 的调查题目，在时间点 $t_3$ 回答有关 $Y$ 的调查题目（各次测量之间的间隔时间要具有一定的意义）。然而，出于实际操作的原因（后勤管理、参与者的流失等等），这一理想状态很少能够实现。其实，许多中介作用检验都应用了依据横截面调查设计所测量得到的相关性数据（即数据是在一个时间点上收集得到的）。而且，在大部分已经发表的包含中介作用分析的论文中，$X$、$M$ 和 $Y$ 的测量都是在同一次调查中完成或是同时进行，或是颠倒次序的（例如，通常情况下是先测量 $Y$ 再测量 $M$，目的是保证因变量的测量不被其他测量所干扰）。当全部测量题目或量表出现在同一次调查中时，研究者就需要从逻辑上或理论上来论证先后次序关系。

在一些情况下——假如先后次序不是很关键——这种论证或许站得住脚。例如，如果 $X$ 是一个稳定的人口学或特征变量，就可以在调查结束时才对它进行测量，这样做不会产生弄错因果作用的顾虑。也就是说，在这种情况下，态度测量（$M$）或行为测量（$Y$）不会对一个早已存在的状态测量（$X$）产生影响。[4] 然而，当 $X$、$M$ 和 $Y$ 测量的都是态度，并且不按先后次序进行时，就会增加在概念上论证作用关系的难度。相反，仅仅满足次序性要求也是不够的。也就是说，在时间点 $t_1$ 测量 $X$，在时间点 $t_2$ 测量 $M$，在时间点 $t_3$ 测量 $Y$，仍然无法保证 $X \rightarrow M \rightarrow Y$ 这一因果或时间顺序一定成立。例如，如果 $X$ 表示态度，$M$ 表示行为倾向，$Y$ 表示个体特质或人口学特征，显然不能将因果顺序总结为 $X \rightarrow M \rightarrow Y$，因为在逻辑上，$Y$ 几乎肯定是先于 $X$ 或 $M$ 而存在的。即使再复杂的统计工具，也无法克服逻辑上的问题。

再者，许多双向关系的存在也可能导致难以确定因果关系或时间顺序的方向。即使绝大部分的变化是沿 $X \rightarrow Y$ 这一方向发生，也会经常发现一些反向效应，即 $X \leftarrow Y$。例如，市场营销人员相信，广告投放会增加销售量（即广告→销售），但广告经费预算却是由上一年的销售情况所决定的（即销售→广告）。

最后要提一下，在许多已经发表的包含中介作用分析的论文中，$M$ 只是对

实验干预因素 $X$ 是否起操控作用的一个校验或核对。这种校验或核对当然不可避免地随操控因素的变化而变化，但没有人会认为，这种以 $X$ 为一种概念建构而将 $M$ 仅仅作为 $X$ 的一个测量的中介模型会对研究文献有什么理论贡献。这种形式的论证也不能表明中介作用的存在，因为它是由概念化过程本身所决定的。[5]

### 排除其他可能的解释

如果已经存在众多的竞争性解释，那么其中的任何一种都会威胁因果关系结论的内部效度，也就是说，会威胁研究者在将观察到的数据模式归因于有待检验的理论观点时的确定性。例如，如果市场营销人员试图将近期的销售增长归因于前期的促销活动，就必须排除另一个可能的原因，即销售的自然性或季节性增长(Churchill & Iacobucci, 2005)。在此我们再一次说明，精心设计的实验方法被认为能够最有效地"排除替代性的因果解释"(Baumrind, 1983:1290)。

排除替代性的解释的考虑既简单又复杂：简单是因为大家都知道这一点，复杂是因为我们往往不清楚是否真正实现了这一目标。例如，McKim 和 Turner(1997)就指出，要排除潜在相关变量的影响是相当困难的，部分原因在于我们无法知道所有可能的原因。Mackie(1974、1980)提出的 INUS 条件认为，我们或许最多只能识别这样一种双变量关系，即第一个变量是某个条件的不充分但必要因素，而该条件却是第二个变量的非必要但充分因素。我们经常碰到的例子是这样一个问题："我们能否弄清楚，是不是电线短路导致了火灾?"(参见 Suppes, 1970)。就 INUS 条件的后半部分("非必要但充分")来说，可能导致火灾的条件有好几种，因此，电线短路本身是一个非必要(一根火柴也可能导致火灾)但却充分的(电线短路当然能够导致火灾)因素。就 INUS 条件的前半部分("不充分但必要")来说，仅仅发生电线短路是一个不充分但必要的因素——即其他因素也必须具备，例如发生电线短路的地方附近存放着可燃物质。INUS 条件就是试图将这些情况都考虑进来以排除竞争性的解释。

这种逻辑上和经验上的因果关系检验有时候会面临更大的挑战，例如当碰

到像 $X \to M_1 \to M_2 \to Y$ 这种延展性的关系链时，或者当碰到像 $X \to M_1 \to Y$ 并且 $X \to M_2 \to Y$ 这种多元中介作用路径时（Karson & Fisher，2005；Salmon，1997）。这里需要再次说明，由于存在前文所说的测量误差和抽样误差的问题，这些变量之间的关系是概率性的。也就是说，研究者需要证实，中介变量在有自变量时，出现的概率要大于没有该自变量时出现的概率，即 $P(M \mid X) > P(M \mid \sim X)$；类似的关系也存在于中介变量与因变量之间，即 $P(Y \mid M) > P(Y \mid \sim M)$，或者像 Sosa 和 Tooley（1993）对完整关系链所作出的条件表述，即如果有 $X \to M \to Y$，则会有 $1.0 > P(Y \mid M) > P(Y \mid X) > P(Y) > 0.0$。

## 中介关系、相关关系与因果关系

诚然，在确立因果关系时所面临的上述问题并不是什么新鲜事。但借助下面的方法，或许可以使这些早已明确的问题显得更为清楚，即阐明 3 个关键概念或变量（$X$、$M$、$Y$）之间的相关矩阵如何产生好几个可以相互替代的竞争性模型，但却同时得出与图 1.1 所示的中介作用结构相一致的结果（参见 Breckler，1990；Lee & Hershberger，1990；MacCallum，Wegener，Uchino & Fabrigar，1993）。我们稍后会详细讨论模型识别与模型等价性的问题（例如，在什么情况下，将无法区分 $X \to M$ 与 $X \leftarrow M$），但现在，我们可以从 Glymour（2001：33）所举的例子入手，先来谈谈这个问题。这个例子是说，如果我们发现控制了 $M$ 之后，变量 $X$ 与变量 $Y$ 之间不相关，确切地说是排除了 $M$ 的影响之后，$X$ 与 $Y$ 之间的偏相关为 0（$r_{XY \cdot M} = 0$），那么这一结果同时意味着 3 个模型：(1)一般意义上的中介作用模型，即 $X \to M \to Y$；(2)整个作用方向倒转的模型，即 $X \leftarrow M \leftarrow Y$；(3)$M$ 同时影响了 $X$ 和 $Y$ 的模型，即 $X \leftarrow M \to Y$。

偏相关为中介作用检验的目标提供了一个简单准确的语言表达。以 $r_{XY \cdot M}$ 这一偏相关为例，它表示在考察 $X$ 与 $M$ 及 $Y$ 与 $M$ 之间的关系时，控制了 $M$ 之后，$X$ 与 $Y$ 之间的线性相关程度。具体来说，如果 $r_{XY \cdot M} = 0$，就表示一旦排除了 $M$ 的作用，$X$ 与 $Y$ 之间就不存在线性关系。Simon（1957：42—43）曾经指出，如果我们试图检验或证实 $X \to M \to Y$ 这一模型，那么就应该有 $r_{XY \cdot M} = 0$ 或者 $r_{XY} = (r_{XM})(r_{MY})$（这两种结果是等价的）。由于数据在本质上只是概率性的，$r_{XY \cdot M}$的

数值或许不一定等于0,所以我们可以换一种说法,即期望$r_{XY\cdot M}$不显著,也就是说,其数值与0之间的差异不具有统计显著性。如果数据同时支持直接和间接作用路径,那么偏相关$r_{MY\cdot X}$的数值虽然会比较小,但应该不会为0或者不具有统计显著性(Asher, 1983:21)。

我们会在本书随后的章节中继续讨论竞争性解释和模型的问题。但最后再提一下,根据我们对学术期刊论文的观察,很少有作者会提到竞争性理论的问题,更别说检验它了。包含中介作用分析的论文在前言部分也很少提到研究的内容之一是要检验 $X$ 影响 $Y$ 的可能方式,也很少明确说明会检验的中介作用是否存在(虽然这样说或许有些吹毛求疵了)。诚然,作为一本书的介绍性章节,或许不必预先提及这些书将会涉及的问题,但是,既然本书旨在分析中介作用,统计检验就必然要涉及,不管是从理论还是从分析的角度来说,都应该予以考量。这种事后考虑提出了一个概念上的存疑,即中介作用的检验究竟是否必要?当然这也提出了对概率性统计检验指标的关注。

# 第 2 章 | 中介作用分析的基本方法

本章将介绍检验中介作用及其经验证据的方法。我们将首先回顾传统的检验中介作用模式的方法，即在回归分析的基础上使用 Baron 和 Kenny(1986)以及 Sobel(1982)的 z 检验。根据这些方法所设定的准则，研究者可以声称中介作用在他们所使用的数据中是“完全的”、“部分的”或是“不存在的”。我们将讨论如何正确地运用这些技术和方法、如何避免易犯的错误以及文献中存在的争议。

随后，我们将介绍另一种方法的可行性，即用有向非循环图来检验中介作用结构。虽然这种方法有助于概念层面的理解，但从经验层面来讲，第三种检验中介作用及其众多扩展形式的方法——结构方程模型(SEM)——将会被证明更具优势。因此，在简要回顾了回归方法和图示方法之后，本书的其他部分都在 SEM 的基础上进行分析。

## 经典的中介作用检验

最流行的检验中介作用的方法是由 Baron 和 Kenny(1986)提出的。根据这一方法，研究者需拟合 3 个回归方程：

$$M = \beta_1 + aX + \varepsilon_1 \quad [2.1]$$

$$Y = \beta_2 + cX + \varepsilon_2 \quad [2.2]$$

$$Y = \beta_3 + c'X + bM + \varepsilon_3 \quad [2.3]$$

方程中的 $\beta$ 表示截距(一般忽略不计)，$\varepsilon$ 表示模型的误差项(一般也忽略)，$a$、$b$、

$c$、$c'$表示回归系数，反映了3个关键变量之间的关系。如果满足下列条件，则认为中介作用可能存在：(1)方程2.1中的$a$显著，也就是说，有证据表明自变量($X$)与中介变量($M$)之间存在线性关系；(2)方程2.2中的回归系数$c$显著，也就是说，自变量($X$)与因变量($Y$)之间存在线性关系；[6](3)方程2.3中的$b$显著，意味着中介变量($M$)有助于预测因变量($Y$)，并且，显示自变量($X$)对因变量($Y$)的直接作用的$c'$与方程2.2中的$c$相比，数值显著变小；(4)方程2.2中的$c$与方程2.3中的$c'$的数值大小的比较是通过$z$检验来判定的(Sobel, 1982)。[7]可以证明，检验$c$(直接作用)与$c'$(控制了间接的中介作用之后的直接作用)的差异与检验中介作用路径($a\times b$)的强度是否大于0是等价的：

$$z=\frac{a\times b}{\sqrt{b^2s_a^2+a^2s_b^2}} \qquad [2.4]$$

方程中$a$和$s_a^2$的值可从方程2.1中得到，$b$和$s_b^2$的值可从方程2.3中得到。[8]

请注意，计算时可使用回归系数，也可使用标准化回归系数，只要它们各自的标准误可得。因为有$SE(\beta_X)=(s_X/s_Y)SE(b_X)$和$SE(\beta_M)=(s_M/s_Y)SE(b_M)$或者$SE(b_X)=(s_Y/s_X)SE(\beta_X)$和$SE(b_M)=(s_Y/s_M)SE(\beta_M)$。

如果$a$或者$b$不显著(或者两者都不显著)，就不存在中介作用，研究者可以总结说$X$对$Y$的影响都是直接的，而不是通过$M$这一中介实现的，那么分析就到此结束。如果满足上述3个条件，研究者就可以总结说至少存在“部分的”中介作用，即$X$对$Y$的影响部分是直接的，部分是间接通过$M$这一中介实现的。如果上述3个条件都得到满足，但$c'$不显著，就可以总结说存在一种“完美的”或“完全的”中介作用，即$X$对$Y$的影响全部通过$M$这一中介间接实现的。[9]

研究论文经常较少报告其分析细节，这种策略确实会“保护”研究者较少受到质疑，撇开这一点不谈，我们还是会发现，许多关于中介作用的分析其实并不正确。例如，虽然大多数论文的分析都严格遵循了上述3个步骤，但并非所有的论文都是如此。而且，许多研究都并没有报告$z$检验的结果。[10]在那些严格履行了所有检验并且完整报告其结果的论文中，大多数结论都存在

"部分的"中介作用。尽管这一基本结论看似合理，即 $X$ 对 $Y$ 的影响部分是直接的，部分是间接的（通过 $M$ 实现的），但它也只不过是一个不怎么令人激动的结果而已。

Baron 和 Kenny(1986)的论文在引导研究者思考中介作用以及为研究者提供一套方法来发现数据中的中介作用模式两个方面影响很大。[11]他们所提出的技术和方法不仅吸引研究者将其应用于各自的研究领域和理论问题，而且催生了方法本身的研究，即促使研究者考察关于中介作用的经验性检验方法的一系列问题。[12]例如，MacKinnon 和他的同事进行了一些方法上的检验，以此比较不同的统计量在发现中介作用模式上的相对优劣以及不同的指标在反映数据中存在中介作用结构强度上的相对功效（MacKinnon et al.，1995、2002）。[13]

## 另一种方法：有向非循环图

关于如何检验和考察因果关系的文献很多，这是科学哲学、统计学和机器学习等学科交叉发展的结果（参见 Hayduk et al.，2003；McKim & Turner，1997；Pearl，2000；Salmon，1997）。学者重新思考因果概念的逻辑以及检验和测定经验关系的方法（例如，Shafer，1996；Spirtes，Glymour & Scheines，2001）。考虑到对复杂性的重新认识，一些学者提到了确立因果关系的"艺术性"（参见 McDonald，2002；Pearl，2000），其中许多人所使用的工具之一就是"有向非循环图"，或简写为 DAG，它与中介作用问题有着十分密切的联系。

"有向非循环图"方法中的术语是这样定义的：该图由表示关系的连线（例如，因果关系或相关关系）和节点（例如，变量 $X$、$M$ 和 $Y$）组成。当图中的连线从一个节点（例如，$X$）出发并指向另一个节点（例如，$Y$）时，就成了"有向图"。如果图中的连线不是循环的（对每一个中介作用路径而言）或者一个因果关系链的终端变量不会反向作用于它的起始变量，"有向图"就成为了"有向非循环图"。也就是说，在"有向非循环图"中，一个表示关系的连线不可能再绕一圈返回到它的起始节点上（循环图则相反，例如：$X \rightarrow M \rightarrow Y \rightarrow X$ 或者 $X \rightarrow M \rightarrow Y \rightarrow M$）。

如果 $X$ 与 $Y$ 之间的关系全部是中介性的，那么 DAG 分析可以总结说，控制了 $M$ 之后，$X$ 与 $Y$ 是相互独立的。DAG 的标注方法会显示，给定 $Y$ 和 $M$ 之后，$X$ 的概率会等于给定 $M$ 之后 $X$ 的条件概率。具体的标注方法如下：若 $P(X \mid Y, M) = P(X \mid M)$，则 $X \perp Y \mid M$。在经验研究中，这些概率可以通过计算偏相关系数 $r_{XY \cdot M}$ 来估计。如果这一偏相关系数不显著（即系数值与 0 之间的差异不具有统计显著性），那么（$X$ 与 $Y$ 之间的）相互独立关系就成立。可以推测，如果 $P(X \mid Y, M) \neq P(X \mid M)$，那么存在中介作用的强度就会减弱（"部分的中介作用"或"不存在中介作用"，但无法区分究竟是哪一种情况）。

DAG 其实对应于某一特定的理论模型，也就是说，$X \rightarrow M \rightarrow Y$ 这一中介作用模式肯定构成了一个有向非循环图。然而，McKim 和 Turner(1997)指出，DAG 并没有回答如何从相关性数据（例如，只能观察到是否相关）推导出因果关系这一问题；Humphreys 和 Freedman(1996)认为，DAG 并没有揭示因果关系结构，却假定了因果关系的存在。

另外，DAG 对于用经验数据分析中介作用来说并不是那么有用，其实，坦白地讲，这一方法本来也不是特地服务于这一目的的。例如，有人设计了 DAG 软件，但其目的只是发现和展示数据中的 DAG(Haughton, Kamis & Scholten, 2006; Shipley, 2000)。在包括许多节点和作用关系的复杂网络中，查找 DAG 会比较有用，因为这些关系网中的作用模式很难一眼辨别，而 DAG 标识出了递归模式的关系链，并将其与循环作用模式进行了区分。然而，这一工作与中介作用分析并没有什么关系，因为在后一种情况下，我们其实知道哪里会有 DAG（而且，中介作用分析中的节点通常很少，一般来说只有 3 个）。确实，某些 DAG 分析技术（例如，"D-separation"关注模型中几个特定的作用路径）类似于估计偏相关，但它们与结构方程模型相比，并不具备方法上的优势(Hayduk et al., 2003)。例如，Pearl(2001)讨论了如何将 $X \rightarrow M \rightarrow Y$ 这一关系链分解为直接作用和间接作用，但结构方程模型也可以做到这一点。通用的计算软件包都没有专门针对基本中介作用模型的拓展情况作程序设计，也没有反映拟合性的统计指标。因此，DAG 并没有为具有中介作用结构的数据提供一个最佳的分析方法。

## 另一个更好的方法：结构方程模型

中介作用研究文献在方法上最大的发展就是学者逐渐用SEM来检验中介作用模式。回归方法操作起来当然容易，Baron 和 Kenny(1986)的论文对此已经讲得很清楚，但研究者经常会遇到更加复杂的情形，需要进一步的指导。另外，SEM本身发展到现在，已被普遍使用，该工具对检验数据中是否存在中介作用模式显得尤为适合。

例如，试想一下这种情形，即研究者分别用多个指标来测量 $X$、$M$ 和/或 $Y$ 这3个概念建构(Baron 和 Kenny 1986年的论文也提到过这种情形，但没有全面地加以讨论)。这时，SEM就可以同时提供一个测量模型和一个结构模型，并将两者完美地结合在一起：首先，SEM提高了概念测量的可靠性；其次，SEM检验了这些概念之间可能存在的中介作用关系。

我们还可以试想一下另外一种情形，即 $X$、$M$ 和 $Y$ 这3个概念嵌套于一个更大的分析网中，它囊括了一些其他原因变量和结果变量。在涉及其他概念或变量时，要检验最为关心的 $X \rightarrow M \rightarrow Y$ 这一中介作用，典型的回归方法在某种程度上有些无从下手。例如，当($X$、$M$ 和 $Y$ 之外的)其他概念或变量作为解释因素时，它们应该被纳入方程2.1至方程2.3中的哪个或哪些方程？它们什么时候又应该被作为因变量来处理？应该比较哪些参数？

一种特殊的扩展性分析网会在研究者检验多元中介作用路径时，或者当因果关系链拉长时出现，前一种情况下的例子如 $X \rightarrow M_1 \rightarrow Y$ 和 $X \rightarrow M_2 \rightarrow Y$，后一种情况下的例子如 $X \rightarrow M_1 \rightarrow M_2 \rightarrow Y$。如何用回归方法的拓展形式来回答这些问题并不明确，但SEM的拓展方法却相当简单和直接，正如我们即将演示的那样。

因此，本书将阐明用SEM来检验中介作用的方法，首先介绍最简单的中介作用结构，其次加入一个测量模型，随后再加入更复杂的结构模型。我们也将使用这些技术和方法来重新探讨其他议题，包括如何检验中等程度的中介作用等。

# 第3章 用结构方程模型来分析中介作用：测量模型

SEM是一个强有力的统计方法，代表了目前最先进的技术发展水平，它使得研究者可以在一系列更广泛的情境中研究中介关系。结构方程模型整合了测量模型和结构模型。我们在本章考察的是结构方程模型的测量部分。我们将考虑中介模型的扩展形式，以纳入任一概念来建构所具有的多重指标。在下一章中，我们考察的是结构方程模型的结构部分。届时我们才会检视实质性的模型延展，即同时考虑不限于 $X$、$M$ 和 $Y$ 的更多的概念建构。

## 多重测量情况下的中介作用

用于中介分析的社会科学数据通常都是由受访者报告的，因此，报告过程中所产生的测量误差将会减弱或冲淡对统计关系的估计。Baron 和 Kenny (1986：1177)承认，他们的(中介分析)基本方法与任何回归分析一样，无法对测量误差作出某种特别的考虑，而只能简单地将其放入总误差项中，从而导致模型拟合上的欠缺，即 $1-R^2$。虽然将单一变量作为某一概念建构的唯一指标在统计上确实是妥当的，但大多数社会科学家却认为用多项量表更可取，这与古典检验理论在哲学认识上是一致的，与越多的测量项目构成了越强的测量工具这一有关信度的理念也是一致的。[14] 1986 年的回归方法仅仅适用于 3 个变量所构成的系统，也就是说，3 个概念建构中的每一个都是由一个指标来测量。然而，图 3.1 却显示了另一种典型的研究情境，即每一个关键的概念建构都由多项量表来测量。这只是一个示例，对于我们将要描述的分析技术来说，测量 $X$、$M$

和 $Y$ 的指标数目可能是两个或以上。

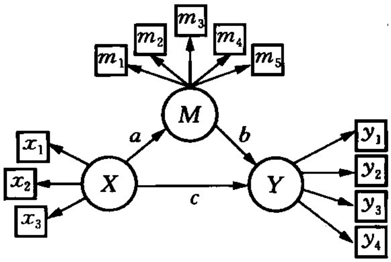



**图 3.1 用多项量表来测量 3 个概念建构**

或许最容易想象的情形是存在多个预测变量 $X_1$、$X_2$、$X_3$，一个中介变量 $M$ 和一个因变量 $Y$。在回归方法中，通过变量替换，方程 2.1 至方程 2.3 会变为：

$$M = \beta_4 + a_1X_1 + a_2X_2 + a_3X_3 + \varepsilon_4 \qquad [3.1]$$

$$Y = \beta_5 + c_1X_1 + c_2X_2 + c_3X_3 + \varepsilon_5 \qquad [3.2]$$

$$Y = \beta_6 + c'_1X_1 + c'_2X_2 + c'_3X_3 + bM + \varepsilon_6 \qquad [3.3]$$

然而，即使是在这种简单的扩展形式中，有两项任务已经变得十分复杂：首先，为了评估中介作用的程度，应该比较哪些回归系数并不清楚；其次，$X_1$、$X_2$、$X_3$ 的作用会因它们之间的多重共性线而发生衰减，这是当预测变量代表了同一个概念建构的多重指标时所固有的问题。

若中介变量 $M$ 或因变量 $Y$ 也应该被一个以上的指标进行测量，情况会变得更复杂。首先，可以看到的是，新的复杂性可能简单地与 $X$ 的复杂性类似。然而，由于 $M$ 和 $Y$ 都在中介作用的一系列回归中充当了因变量的角色，因此，$M$ 和 $Y$ 数量的增加会要求产生更多的预测方程。例如，在 3 个中介变量和 2 个因变量的情况下，应有如下方程：

$$M_1 = \beta_7 + a_1X_1 + a_2X_2 + a_3X_3 + \varepsilon_7 \qquad [3.4]$$

$$M_2 = \beta_8 + a_1X_1 + a_2X_2 + a_3X_3 + \varepsilon_8 \qquad [3.5]$$

$$M_3 = \beta_9 + a_1X_1 + a_2X_2 + a_3X_3 + \varepsilon_9 \qquad [3.6]$$

$$Y_1 = \beta_{10} + c_1X_1 + c_2X_2 + c_3X_3 + \varepsilon_{10} \quad [3.7]$$

$$Y_2 = \beta_{11} + c_1X_1 + c_2X_2 + c_3X_3 + \varepsilon_{11} \quad [3.8]$$

$$Y_1 = \beta_{12} + c'_1X_1 + c'_2X_2 + c'_3X_3 + bM_1 + \varepsilon_{12} \quad [3.9]$$

$$Y_1 = \beta_{13} + c'_1X_1 + c'_2X_2 + c'_3X_3 + bM_2 + \varepsilon_{13} \quad [3.10]$$

$$Y_1 = \beta_{14} + c'_1X_1 + c'_2X_2 + c'_3X_3 + bM_3 + \varepsilon_{14} \quad [3.11]$$

$$Y_2 = \beta_{15} + c'_1X_1 + c'_2X_2 + c'_3X_3 + bM_1 + \varepsilon_{15} \quad [3.12]$$

$$Y_2 = \beta_{16} + c'_1X_1 + c'_2X_2 + c'_3X_3 + bM_2 + \varepsilon_{16} \quad [3.13]$$

$$Y_2 = \beta_{17} + c'_1X_1 + c'_2X_2 + c'_3X_3 + bM_3 + \varepsilon_{17} \quad [3.14]$$

即使是适度地扩展到3个概念建构都具有若干个测量变量这种情况，方程的数量和检验组合的数量都大大增加。这时进行中介分析有两种选择。首先，大多数研究者会计算所有量表——$X_1$、$X_2$、$X_3$——中各项的均值，得到 $\bar{X}$，如此类推。用这些汇总数值来代表各个概念建构，就会使原本复杂的测量情形得以简化，因此模型就十分类似方程 2.1 至方程 2.3，即：

$$\bar{M} = \beta_{18} + a\bar{X} + \varepsilon_{18} \quad [3.15]$$

$$\bar{Y} = \beta_{19} + c\bar{X} + \varepsilon_{19} \quad [3.16]$$

$$\bar{Y} = \beta_{20} + c'\bar{X} + b\bar{M} + \varepsilon_{20} \quad [3.17]$$

因此，一个多变量问题就被有效地简化成了一个只具有单一预测变量的问题。我们接下来会说明，这种解决方案虽然可以接受，但却并不是最理想的。

## 中介作用与结构方程模型的测量模型

第二种选择是使用结构方程模型。结构方程模型正是被设计出来以解决方程 3.4 至方程 3.14 这样的线性方程系统。SEM 在社会科学中已经司空见惯并有可供运行的软件，其逻辑也已经被广泛地理解(Kline, 1998)。SEM 为检验

概念建构之间的中介关系提供了目前最先进的分析方法，尤其是当多重题项或多重变量被测量以用来捕捉关键概念建构的时候。[15] Brown(1997)和 Long(2006)的著述在这方面是极好的入门读物。

如果 $X$、$M$ 或 $Y$ 中有一个或几个是通过多项量表来进行测量的，则这种情形下的基本的中介作用探索就可以用表 3.1 中的描述来总结。该表呈现了单一测量与多重测量的所有组合形式，并标明了在哪些情形下可以使用回归方法。具体来说，只有当 $X$、$M$ 和 $Y$ 这 3 个概念建构都由单一分值来表示时，就可以用回归方法，这里所指的单一分值，既可以是单一变量(例如，$M$)，也可以是用来测量某一概念建构的多重题项的得分均值(例如，$\{M_1, M_2, M_3\}\to\overline{M}$)。

**表 3.1 具体数据特性下的分析方法:结构方程模型和回归**

| 中介变量($M$)的测量数目[a] | 自变量($X$)的测量数目[a] | 因变量($Y$)的测量数目[a] | | |
|---|---|---|---|---|
| | | 一个因变量测量 | 多个因变量测量 | 一个汇总性的因变量测量 |
| | | $Y$ | $Y_1$、$Y_2$、$Y_3$ | $\overline{Y}$ |
| | $X$ | SEM、Reg[b,c] | SEM | SEM、Reg[c] |
| $M$ | $X_1$、$X_2$、$X_3$ | SEM | SEM | SEM |
| | $\overline{X}$ | SEM、Reg[c] | SEM | SEM、Reg[c] |
| | $X$ | SEM | SEM | SEM |
| $M_1$、$M_2$、$M_3$ | $X_1$、$X_2$、$X_3$ | SEM | SEM | SEM |
| | $\overline{X}$ | SEM | SEM | SEM |
| | $X$ | SEM、Reg[c] | SEM | SEM、Reg[c] |
| $\overline{M}$ | $X_1$、$X_2$、$X_3$ | SEM | SEM | SEM |
| | $\overline{X}$ | SEM、Reg[c] | SEM | SEM、Reg[c] |

注:结构方程模型(SEM)总是适合的。若任一概念建构是用多重题项来测量的(但不是用一个均值来反映)，就不能使用回归方法(Reg)。

a. 表中描述的是测量数目为 3 的情况(例如，$M_1$、$M_2$、$M_3$)，需着重指出的是，本书所阐释的多项测量原则也适用于两个或更多题项的情况。

b. 这种情形描绘了经典的中介分析，即每一个概念建构 $X$、$M$ 和 $Y$ 都只有一个测量。

c. 有这一标记的 8 个格子表示建模时只有一个测量(例如，$M$)或衡量(例如，$\overline{M}$)。

然而，更为一般化的结论是，SEM 可以在任意一种情形下使用，这也是唯一一项适用于大多数研究情境的分析技术。即使是在最简单的数据情形下(回归在此种情形下看似可作为一种替代性的方法)，我们也会表明，就算回归方法与

SEM之间存在某种一致性,SEM技术无论在理论层面还是经验统计层面都是一种更具优越性的方法。

总而言之,上表表明,结构方程模型总是适用的,而只要任何一个概念建构是由多重题项来测量的(并非用一个均值来反映),就不能使用回归方法。因此,对于分析中介结构来说,在表3.1中的任何情形下,SEM都提供了一个自然的选择和一项最适合的技术。

## 实例阐释

我们刚才已经论证了SEM方法在统计上和理论上都优于回归方法。在本节中,我们用几个说明性的数据和分析来展示SEM的这种优越性。首先,让我们来比较结构方程模型和回归方法在最简单的数据情形下的表现,即只有一个$X$变量、一个$M$变量和一个$Y$变量来分别代表3个概念建构。在这种(也只有在这种)最简单的情形下,我们将看到SEM和回归方法这两种分析技术会得到一些相似的结果,但我们也将发现它们之间的不同。

从相关系数推导出路径系数的方程如下(Asher, 1983; James et al., 1982):

$$a = r_{XM}$$
$$b = \left(\frac{1}{1-r_{XM}^2}\right)(r_{MY} - r_{XM}r_{XY}) \qquad [3.18]$$
$$c = \left(\frac{1}{1-r_{XM}^2}\right)(r_{XY} - r_{XM}r_{MY})$$

与之对应,从路径系数推导出相关系数的方程如下:

$$r_{XM} = a$$
$$r_{XY} = c + ab \qquad [3.19]$$
$$r_{MY} = b + ac$$

例如,一位研究者得到这样一个相关矩阵:$r_{XM} = 0.55$, $r_{XY} = 0.45$, $r_{MY} = 0.63$。根据这些相关系数,我们可以用方程3.18来估计路径系数$a$、$b$、$c$(它们

分别代表了 $X \to M$、$M \to Y$、$X \to Y$)，由此得到 $a=0.55$，$b=0.55$，$c=0.15$。这些路径系数表明，由 $X$ 所解释的 $Y$ 的变异中，有 2/3 是通过中介变量 $M$ 间接实现的，即：

$$\frac{\hat{a} \times \hat{b}}{(\hat{a} \times \hat{b})+\hat{c}}=\frac{0.55 \times 0.55}{(0.55 \times 0.55)+0.15}=66.9\%$$

在这种最简单的数据情形下，即只有 3 个概念建构并且每一个概念建构都只有一个测量来表示，无论用回归方法还是用 SEM 来计算，所估计的路径系数都是相等的。两种分析技术都是以方程 3.18 为基础来对参数作出估计的。

但是，两种分析技术也存在一个重要的系统性差别。回归方法中的系数标准误要比 SEM 中的大一些。对一个不算大的样本 $n=50$ 来说，用回归方法得到的标准误是：$SE(a)=0.121$，$SE(b)=0.134$，$SE(c)=0.134$；用 SEM 得到的标准误是：$SE(a)=0.119$，$SE(b)=0.131$，$SE(c)=0.131$。两者之间的差异虽然小(一般来说，这种差异会随着样本的增大而减小)，但却是系统性的。较小的标准误在统计上是有利的，由于它意味着更高的估计精确性，因而也是更好的。

SEM 相对于回归方法的优越性在于它会减小标准误，因为结构方程模型中所有的参数是被同时估计的。分别拟合 3 个回归模型会导致偏误，其原因是存在缺失变量。相反，在一个模型中同时对所有关系进行建模，在统计上总是优于以零散的形式分别进行建模，因为(对某种关系的)估计能够在统计上控制或者排除其他关系(的影响)。因此，上述经验分析结果并非某种巧合，也并非因为数据示例具有特殊性，相反，它是由统计理论所决定的，这种统计理论即指同时拟合的方程会得到更一致的估计。不管在理论上还是在经验上，拟合一个 SEM 模型都会比拟合 3 个回归方程得到更有效、更精确的估计。同理，既然 SEM 最大似然法与普通最小二乘回归相比会得到更小的标准误，那么，检验中介作用存在与否的 $z$ 检验"在一致性上最具效力"的这种说法就更加适用于 SEM 而非回归方法，这一点不足为奇。

在同时拟合一个模型时，对所有参数和标准误(它们构成了 $z$ 检验的组成部分)的估计都是以模型中所呈现的同一批影响效应为条件的。对 $X \to M$、$X \to Y$、$M \to Y$ 每一种影响效应的估计都在同样的建模情境下进行，而在回归

方法建模中，在对 $X \to M$ 和 $X \to Y$ 的影响效应首先进行估计的模型中是不存在其他影响效应的，这使得每一种影响效应既无法与其他任何一种影响效应比较，也无法与全模型 $\{X, M\} \to Y$ 进行比较。目前，z 检验都是在估计了不同的回归模型之后得出的，而这些回归模型的估计却是以预测变量的不同组合为条件的。简言之，回归方法比较的是苹果和柑橘，而结构方程模型则（真正）比较了苹果和苹果。

在（寻求）较小的标准误的驱使下，最近对 SEM 和回归分析方法进行比较的研究都用到了一系列蒙特卡罗模拟（Iacobucci, Saldanha & Deng, 2007）。模拟研究的长处在于，总体参数是建构的和已知的，因而研究者对总体中的真实关系有所把握，故其逻辑就是研究回归方法或结构方程模型能够在多大程度上正确地发现和复原这种关系模式。这些研究的结果都表明，对各种规模的样本来说，SEM 都比回归方法更有效力，尤其是在小样本（$n = 30$）的情况下，两种方法之间的差异最大，研究者会更偏向于 SEM，因为他们可从额外的补偿性检验能力中获益。SEM 方法在发现概念建构之间的所有相关模式所构成的中介结构上也具有优越性，并且当研究者特别希望发现数据中存在较强的中介效应时（例如，75%和 100%的中介作用），其优势最大。[16]

因此，即使是在这种最简单的数据情形下，即在只有 3 个概念建构并且每一个概念建构都只有一个测量这种典型情况下，也会存在对回归方法和结构方程模型进行选择的问题。而且，SEM 无疑是一种更好的分析技术。基于同步估计的精确性，SEM 的分析结果会在许多方面有利于研究者，包括它更有可能发现存在的中介作用模式，更接近总体的、已知的结构性特征，更具有统计上的可靠性。最后，SEM 允许研究者在难以或无法运用回归方法的情况下，可以考察中介作用，我们现在就将展示这一点。

## 拟合测量模型

下面，我们来考虑一下对数据拟合测量模型时的一些概念性和实际性的议题。首先讨论三变量模型，其次引入多重测量，之后提供 LISREL 语句并对其

进行解释。

### 通过 SEM 拟合基本的三变量中介结构

当使用 SEM 这一工具来分析中介作用时，只有一个模型会被拟合。对 $X \to M$、$X \to Y$、$M \to Y$ 这些作用路径的估计会同时进行。研究者并不是用回归方法来拟合一系列的方程或模型。这个结构方程模型遵循：

$$Y = \Gamma X + BY + \Psi \qquad [3.20]$$

在方程 3.20 中，$X$ 被称为"外生变量"，而 $Y$ 则同时包括了 $M$ 和 $Y$，它们被称为"内生变量"（基本上，$X$ 没有预测变量，但 $M$ 和 $Y$ 都有一个或几个预测变量）。中介分析的主要兴趣在于结构参数 Γ 和 B。这些矩阵有如下形式：

$$\Gamma = \begin{matrix} & X \\ M & \\ Y & \end{matrix} \begin{bmatrix} \gamma_{MX} \\ \gamma_{YX} \end{bmatrix} \to B = \begin{matrix} & M & Y \\ M & & \\ Y & & \end{matrix} \begin{bmatrix} \beta_{MM} = 0 & \beta_{MY} = 0 \\ \beta_{YM} & \beta_{YY} = 0 \end{bmatrix}$$

在矩阵 Γ 中，外生变量（或概念建构）构成了列，它对行上的内生变量 $M$ 和 $Y$ 的作用分别是 $r_{MX}$ 和 $r_{YX}$。[17]矩阵 B 大部分由 0 构成（对应于没有被估计的作用路径，例如，$Y \to M$），但 $\beta_{YM}$ 项需要估计，它代表了 $M \to Y$ 这一作用路径。我们很快就会详细解释矩阵 Γ 和 B 中的各项，但为使这里的论述更加完整，我们还需指出，矩阵 Ψ 也要被估计，它相当于内生模型在拟合方面的欠缺性，即试图预测 $M$ 和 $Y$ 时 $1 - R^2$ 的值：

$$\Psi = \begin{matrix} & M & Y \\ M & & \\ Y & & \end{matrix} \begin{bmatrix} \Psi_M & 0 \\ 0 & \Psi_Y \end{bmatrix}$$

回到最为重要的结构参数，如果存在一个显著的 $X \to Y$ 作用路径（即 $r_{YX}$ 的估计值），就意味着 $Y$ 的变异中至少有一部分可以由 $Y$ 与 $X$ 之间的直接关系来解释，并且，纳入一个中介因素本身并不能充分支持 $X$ 影响 $Y$ 这一过程（事实上，这一中介作用路径可能并不必要）。如果 $X \to M$ 和 $M \to Y$ 两个作用路径（即

$r_{MX}$和$\beta_{YM}$)都显著,那么$Y$的变异中至少有一部分可以由$Y$与$X$之间经由$M$发生的间接或中介关系来解释。如果这两个间接作用路径($X \rightarrow M$或$M \rightarrow Y$)中只有一个显著,那么中介作用逻辑可能会留待以后的研究来确定(例如,不显著的那个作用路径可能使用了不可靠的测量工具,或者研究所使用的样本规模不够大),但当前的研究结果并不能支持存在中介作用这一结论。

如果$X \rightarrow M$和$M \rightarrow Y$两个路径系数都显著,研究者就应该进行后续的z检验。这一后续检验在SEM方法中仍然重要,因为不管$X \rightarrow Y$是否显著,问题在于评估中介作用相对于直接作用的大小。假如$X \rightarrow M$的估计值为0.7,$M \rightarrow Y$的估计值为0.6,$X \rightarrow Y$的估计值为0.3,再假定前两项作用显著而后一项作用不显著。从本质上来讲,这可能支持了中介作用而非直接作用假设。然而,上述显著性检验只是表明了0.7和0.6都分别大于0,0.3并不大于0。但是,间接作用可能并不显著地大于直接作用。也就是说,0.7可能并不显著地大于0.3,与之相应,间接作用的乘积$0.7 \times 0.6$也可能并不大于0.3。而z检验就直接比较了间接作用($0.7 \times 0.6$)与直接作用(0.3)的相对大小。

虽然z检验的形式与方程2.4相同,但SEM的优势在于,该检验中的所有成分——包括系数$a$、$b$以及它们各自的标准误$s_a$、$s_b$——都来自同一个模型,即方程3.20。由于每一个估计都是在同样的基础上,即排除了所有一系列类似的其他效应后得到的,它们就在最大程度上实现了无偏性。

为使本书对那些希望使用结构方程模型来检验中介作用的读者更具实用性,我们将会把这些统计和矩阵概念转换为软件指令。我们现在就来呈现运行上文提出的每一个模型的LISREL语句,因为LISREL可能是最流行的SEM软件包。[18]在本书的附录部分,我们提供了用其他流行的软件包(SAS的PROCCALIS、EQS和AMOS)来拟合基本中介模型的语句。我们虽然会对这些语句进行解释,但却只限于简单的解释,读者需要对LISREL有一定的了解(Jöreskog & Sörbom, 1997)。我们从运行最简单的三变量模型入手来解释LISREL语句(一个三变量模型已经涵盖了已有文献中的大多数中介分析,但我们希望在本书阐述了若干种扩展形式之后,能够在将来的研究中见到更复杂的分析和推理)。

在下面的指令中，第一行以关键词"Title"开始，以句点结束。第二行说明了数据的结构("da")。在本例中，该行语句表明输入变量的数目("ni")为 3，观察个案的数目("no")或样本规模为 100，要分析的矩阵("ma")是协方差矩阵("cm")。关键词"la"表示变量标记见下一行。关键词"cm"表示协方差矩阵是以对称("sy")形式输入的。关键词"se"允许用户对变量进行选择，或者是对变量进行重新排序以使内生变量(在一个三变量中介模型中即 $M$ 和 $Y$)首先出现。

模型表述("mo")行意味着，本例是一个不存在多重测量的路径模型，因为 $Y$ 或内生变量(即 $M$ 和 $Y$)的数目与内生(eta 和 $\eta$)概念建构的数目都为 2。同样，$X$ 的数目与外生(ksi 和 $\xi$)概念建构的数目都为 1。因子矩阵 lx 和 ly 被设定("fi")为恒等("id")矩阵(这再一次表明指标变量与潜在变量之间是一一映射的)，基于同样的原因，测量误差矩阵 td 和 te 被设定为 0。我们随后会对 eta、ksi、lx、ly、td、te 矩阵作出更详尽的描述——当中介作用中的基本概念建构是用多个变量来表示时，这些矩阵会更有用。

B 和 Γ 矩阵是由模式("pa")来界定的，它指明哪些作用路径应该被估计(即"1")，哪些则应该被设定为 0(即"0")。软件对这两个矩阵的读取类似于假定一个列变量在影响行变量。因此，B 矩阵中唯一的"1"，即 $\beta_{2,1}$，表明第一个内生变量(根据"se"对变量的列举顺序，即 $M$)对第二个内生变量(即 $Y$)具有影响效应。结果输出是通过指令"ou"完成的。我们鼓励读者去尝试这些语句，看是否能够得到前文所给出的结果，即 $a = 0.55$，$b = 0.55$，$c = 0.15$。

```
Title: Mediation With Three Constructs, One Measure Each.
da ni = 3 no = 100 ma = cm
la
x m y
cm sy
1.00
0.55 1.00
0.45 0.63 1.00
```

```
se
m y x
mo ny = 2 ne = 2 nx = 1 lx = id, fi td = ze, fi ly = id, fi te = ze, fi be =
fu, fr ga = fu, fr
pa ga
1
1
pa
be
0 0
1 0
out me = ml rs ef
```

## 有多重测量时用 SEM 来拟合中介作用并同时纳入测量模型

研究者知道，当用多项式量表来测量 $X$、$M$ 和/或 $Y$ 时，会提高其可靠性，也会有助于厘清这些测量题项所代表的概念建构之间的关系本质。[19]如果出现了多项式量表(例如，用 $X_1$、$X_2$、$X_3$ 来测量自变量 $X$ 这一概念建构)，研究者则不清楚如何应用回归方法来检验中介作用(例如，将 $X_1$、$X_2$、$X_3$ 代入方程 2.1 至方程 2.3 时，会由于多元共线性问题而造成困惑)。

在这种情况下，研究者通常会计算多项式量表的均值，即 $\bar{X}$、$\bar{M}$、$\bar{Y}$，再将其代入类似于方程 2.1 至方程 2.3 的方程体系中，即方程 3.15 至方程 3.17。然而，虽然通过对多重测量取均值(即，$\{X_1, X_2, X_3\} \rightarrow \bar{X}$)来简化分析十分常见，但这种做法并没有充分利用数据的全部优势，相反，用结构方程模型来拟合中介作用结构却可以在一个测量模型中体现这些多重测量题项。如果将 $X$、$M$ 和 $Y$ 的多重测量汇总为均值加以分析，再将其结果与使用一个完整的结构方程模型(即同时包括了测量模型和路径模型)所得到的结果进行比较，或许毫不奇怪的是，后一种结果会更具优势(Iacobucci et al., 2007)。完整的结构方程模型正

确地识别了研究总体的特性,并几乎完全复原了由中介关系所导致的那部分变化,但对均值的分析则高估了中介作用的程度(例如,这种分析会偏向于得出样本中存在部分或完全的中介作用这一结论,即使是研究总体中并不存在任何中介作用时也会如此)。

因此,当用多项式量表测量 $X$、$M$ 或 $Y$ 时,研究者若通过取其平均值而不用 SEM 来进行建模,这其实是对数据的一种损害。在 SEM 方法中,重要的是让数据按照其固有的特性来“说话”——如果某一概念建构有 3 个测量指标,那么对中介作用的检验就应该同时纳入一个正确的、同构的测量模型和一个结构模型,而不是将数据处理为其均值。计算量表的均值可能看似简化了分析过程,但却牺牲了结论在实质上和理论上的准确性,研究者会为这种走捷径的行为付出代价,这种代价就是,他们经常会得出有关中介作用检验的错误结论。

对数据的建模应该以一种能够代表其结构的方式进行,这并不只是一个“理论上的”论断。它也能在实际上对研究者有所帮助。在用一个结构方程模型对多重测量进行建模时,测量误差被从模型误差中分离出来。因此,概念建构之间的关系也被澄清了,并且通常还被强化了。

所以,研究者应该运行一个完整的结构方程模型,而不是将多项式量表汇总为其均值。这个完整的结构方程模型由结构模型和测量模型两部分构成,前者说明了 $X$、$M$ 和 $Y$ 这 3 个概念建构之间的关系本质,后者说明了将这些概念建构映射于各自的指标变量的情形。

我们所说的多重测量题项,是指由 3 个或更多测量题项所组成的量表。在理论上,多重测量也包括由两个题项所组成的量表,但由 3 个或更多题项所组成的量表有助于研究者避免欠识别的问题(除非在理论上或经验上有两个或更多的题项极其相关)。Bagozzi 和 Baumgartner(1994)建议,由 3 到 5 个题项所组成的量表或许是最理想的(也就是说,题项的数目最少要有 3 个,而 5 个以上可能就过犹不及了)。

在多项式量表的情况下,用 LISREL 的标记方法,我们现在来区分 $\xi$ 和 $X$,前者是指外生变量所表示的概念建构,后者则表示测得的外生变量。与之类似,$\eta$ 是指内生的概念建构,由 $M$ 和 $Y$ 的变量来代表。完整的结构方程模型由

如下结构方程：

$$\eta = \Gamma\xi + \mathrm{B}\eta + \Psi \qquad [3.21]$$

和下面两个测量模型(分别代表了将 $\xi$ 映射于 $X$ 和将 $\eta$ 映射于 $M$ 和 $Y$)构成：

$$Y = \Lambda_Y\eta + \theta_\varepsilon \qquad [3.22]$$

$$X = \Lambda_X\xi + \theta_\delta \qquad [3.23]$$

考虑 3 个 $X$ 变量、3 个 $M$ 变量和 2 个因变量 $Y$ 这种情形(即方程 3.4 至方程 3.14)，就会得到方程 3.21 至方程 3.23 中的 SEM 矩阵。结构参数是 $\Gamma$ 和 B，它们与中介分析最相关，也是理论最关心的。由于它们代表了 3 个关键的概念建构，所以其结构与之前提及的只有 3 个测量的路径模型的情形相似：

$$\Gamma = \begin{array}{c} \\ \eta_M \\ \eta_Y \end{array}\begin{array}{c} \xi \\ \begin{bmatrix} \gamma_{MX} \\ \gamma_{YX} \end{bmatrix} \end{array},\quad \mathrm{B} = \begin{array}{c} \\ \eta_M \\ \eta_Y \end{array}\begin{array}{c} \eta_M \qquad\qquad \eta_Y \\ \begin{bmatrix} \beta_{MM} = 0 & \beta_{MY} = 0 \\ \beta_{YM} & \beta_{YY} = 0 \end{bmatrix} \end{array}$$

方程 3.22 和方程 3.23 中的新矩阵 $\Lambda_X$ 和 $\Lambda_Y$ 包含了 $X$ 变量和 $Y$ 变量的因子载荷(遵循标准的处理方式，每个因子的第一个载荷被设定为 1 以实现模型的标准化)。

$$\Lambda_X = \begin{array}{c} \\ X_1 \\ X_2 \\ X_3 \end{array}\begin{array}{c} \xi \\ \begin{bmatrix} \lambda_{X_{11}} = 1 \\ \lambda_{X_{21}} \\ \lambda_{X_{31}} \end{bmatrix} \end{array}$$

$$\Lambda_Y = \begin{array}{c} \\ M_1 \\ M_2 \\ M_3 \\ Y_1 \\ Y_2 \end{array}\begin{array}{c} \eta_M \qquad\qquad \eta_Y \\ \begin{bmatrix} \lambda_{Y_{11}} = 1 & 0 \\ \lambda_{Y_{21}} & 0 \\ \lambda_{Y_{31}} & 0 \\ 0 & \lambda_{Y_{42}} = 1 \\ 0 & \lambda_{Y_{52}} \end{bmatrix} \end{array}$$

另外两个矩阵 $\Theta_\delta$ 和 $\Theta_\varepsilon$ 包含了测量误差(其中,与被设定为 1 的因子载荷相对应的误差项被设定为 0,其他项则会被估计)。

$$\Theta_\delta = \begin{array}{c} X_1 \\ X_2 \\ X_3 \end{array} \overset{\begin{array}{ccc} X_1 & X_2 & X_3 \end{array}}{\begin{bmatrix} \theta_{\delta_{11}}=0 & 0 & 0 \\ 0 & \theta_{\delta_2} & 0 \\ 0 & 0 & \theta_{\delta_3} \end{bmatrix}}$$

$$\Theta_\varepsilon = \begin{array}{c} M_1 \\ M_2 \\ M_3 \\ Y_1 \\ Y_2 \end{array} \overset{\begin{array}{ccccc} M_1 & M_2 & M_3 & Y_1 & Y_2 \end{array}}{\begin{bmatrix} \theta_{\varepsilon_1}=0 & 0 & 0 & 0 & 0 \\ 0 & \theta_{\varepsilon_2} & 0 & 0 & 0 \\ 0 & 0 & \theta_{\varepsilon_3} & 0 & 0 \\ 0 & 0 & 0 & \theta_{\varepsilon_4}=0 & 0 \\ 0 & 0 & 0 & 0 & \theta_{\varepsilon_5} \end{bmatrix}}$$

如前文所述,矩阵 $\Psi$ 包含了内生模型的欠拟合性,类似于 $1-R^2$。而 $\Phi$ 则度量了外生的概念建构(如果外生的概念建构不止一个,则捕捉了它们之间的相关关系)[20]:

$$\Psi = \begin{array}{c} \eta_M \\ \eta_Y \end{array} \overset{\begin{array}{cc} \eta_M & \eta_Y \end{array}}{\begin{bmatrix} \Psi_{\eta_M} & 0 \\ 0 & \Psi_{\eta_Y} \end{bmatrix}}, \quad \Phi = \begin{array}{c} \xi \end{array} \overset{\xi}{[\phi]}$$

## LISREL 语句

对于有 3 个概念建构、$X$ 有 3 个测量指标、$M$ 有 3 个测量指标、$Y$ 有 2 个测量指标的模型,LISREL 语句如下:

```
Title: Mediation With Three Constructs, 3 X's, 3 M's, and 2 Y's.
da ni = 8 no = 100 ma = cm
la
```

```
x1 x2 x3 m1 m2 m3 y1 y2
cm sy
1.00
0.30 1.00
0.30 0.30 1.00
0.30 0.30 0.30 1.00
0.30 0.30 0.30 0.30 1.00
0.30 0.30 0.30 0.30 0.30 1.00
0.30 0.30 0.30 0.30 0.30 0.30 1.00
0.30 0.30 0.30 0.30 0.30 0.30 0.30 1.00
se
m1 m2 m3 y1 y2 x1 x2 x3
mo ny = 5 ne = 2 nx = 3 nk = 1 lx = fu, fr td = di, fr ly = fu, fr te = di, fr
be = fu, fr ga = fu, fr
pa lx
0
1
1
pa ly
0 0
1 0
1 0
0 0
0 1
pa ga
1
1
pa be
```

```
0 0
1 0
va 1.0 lx(1, 1) ly(1, 1) ly(4, 2)
fi td(1) te(1) te(4)
va 0.0 td(1) te(1) te(4)
out me = ml rs ef
```

# 第 4 章 用结构方程模型来分析中介作用：结构模型

结构方程模型由测量模型和结构模型两部分构成。前一章关注的是测量模型，这一章关注的是结构模型。

研究者在试图考察 $X$、$M$ 和 $Y$ 这些概念建构之间的中介作用时，通常都是在一个相对狭窄的视野内进行，也就是说，他们往往只将这 3 个概念建构纳入分析（例如，简短的实验或调查，这是为了实现效率）。然而，即使研究者并不明确地承认存在一个更广的分析网，所有的概念建构之间的关系其实都或多或少地嵌入了一个更大的图景，正如图 4.1 所展示的那样。科学哲学家和方法论者鼓励用这种更广的分析网来对社会现象提出最丰富的观察和解释（Cronbach & Meehl, 1955）。例如，在中介作用分析中，一些研究者可能希望检验比一般情况下更长的中介作用关系链，比如可能是 $X \rightarrow M_1 \rightarrow M_2 \rightarrow Y$。其他研究者可能想探究两个中介系统中的哪一个对因变量的作用更强，例如比较这样两个作用路径：$X \rightarrow M_1 \rightarrow Y$ 和 $X \rightarrow M_2 \rightarrow Y$。

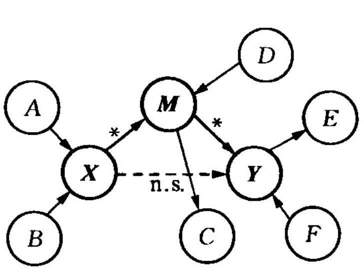



**图 4.1 在一个更广的分析网中检验中介作用**

显然，各种可能性是趋于无穷的。幸运的是，SEM 在允许纳入多重测量之

外，还能够方便地处理超过3个关键概念建构（$X$、$M$和$Y$）情况下的理论归纳和模型扩展。更广的分析网的各种情况都可以被建模，不管新纳入的概念建构构成了原有的3个关键概念建构的原因还是结果。在本章中，我们就来考察一下纳入额外的概念建构时所产生的一些问题，其中的部分已经有了解决方法。

## 概念回顾

对中介作用建模的技术发展来说，至少多加入一个概念建构（以$Q$表示）是尤其必需的，即使研究者更关心$X$、$M$和$Y$而非$Q$(Bentler, 2001)。加入额外概念建构的主要理论目的是提高分析网的复杂性，这使得根据观察到的数据模式提出似乎合理的替代性理论的难度加大，因此保证了结果在统计上和概念上都更为确定。

加入额外概念建构的主要统计目的是改进自由度。中介模型就3个概念建构提出3个连接关系，因此，它具有"恰好识别"的特征。这种特征的其中一个含义是，影响效应的方向在经验上是不确定的，例如，我们不清楚是$X \rightarrow M$还是$M \rightarrow X$（MacCallum et al., 1993; McDonald, 2002; Fienberg, 1981: 122—123）。在前文描述的例子中（即$r_{XM} = 0.55$，$r_{XY} = 0.45$，$r_{MY} = 0.63$），$X \rightarrow M \rightarrow Y$这一中介作用的路径系数分别为$a = 0.55$，$b = 0.55$，$c = 0.15$。如果整个作用方向是倒转的，即$Y \rightarrow M \rightarrow X$，估计值就会出现一定的差异，分别为：0.63($Y \rightarrow M$)，0.44($M \rightarrow X$)，0.17($Y \rightarrow X$)。如果某研究者提出$X \rightarrow Y \rightarrow M$，则估计值会是0.45($X \rightarrow Y$)，0.48($Y \rightarrow M$)，0.33($X \rightarrow M$)。更大的问题是，3个模型（以及更多的模型）的拟合统计量并无二致，都是完美的（例如，Bentler的比较拟合指标CFI都等于1，均方根误差RMSE都等于0）。[21]

理论应该有助于区分可相互替代的模型（这是以参数估计值而非总的拟合统计量为基础，因为前者是可变的而后者则不是），但是，具有竞争性的模型很少被文献所提及，也更少被检验。而且，从同一个理论视角导出的竞争性的因果模型可以具有同样的合理性。例如，如果$X$＝爱好，$M$＝认知，$Y$＝行为，支持$X \rightarrow M$或者$M \rightarrow X$这两种相反理论的模型会同时存在，同理，支持$M \rightarrow Y$或者$Y \rightarrow M$的模型也会同时存在，以此类推（Breckler, 1990）。

使用多项式量表并不能解决这一问题。虽然原始协方差矩阵确实变大了，这也导致自由度的增加，但多余的自由度却是“虚假的”，因为它们是对测量模型的准确性有贡献，而不是对关键的结构模型的自由度有贡献。无论测量了多少个题项，只要有3个概念建构，就有3组概念建构之间的相互关系，并且，如果我们试图在中介分析中估计3个路径，那么模型就是“恰好识别”的。然而，4个概念建构就会有6组相互关系。在这种情况下，如果要估计4个路径，那么就有两个多余的自由度可以用来检验竞争性模型的拟合优越性。我们偏向于这种“过度决定”的情形，因为在这种情形下，未知项的数目要少于系统中已知的方程数目或数据特征，由此得出的每一个估计值都可能会更加确定。

当加入一个新的概念建构 $Q$ 来依次作为 $X$、$M$ 和 $Y$ 的一个原因或结果时，就会相应地多出一个路径，图 4.2 展示了这些路径所有可能的位置。在正式的研究中，$Q$ 的位置会由理论规定，但在这里，我们会考察 $Q$ 的各种不同角色以澄清关于结构方程模型和中介作用的几项要点。图中每个递归的、可识别的结构模型都达到了增加自由度以对模型的拟合性进行评估的目的，也就是说，任何一个模型都不具有人为的完美拟合性。然而，我们必须说明的是，6个模型并不是等价的，其中模型 c 和模型 e 尤其不应该被使用。

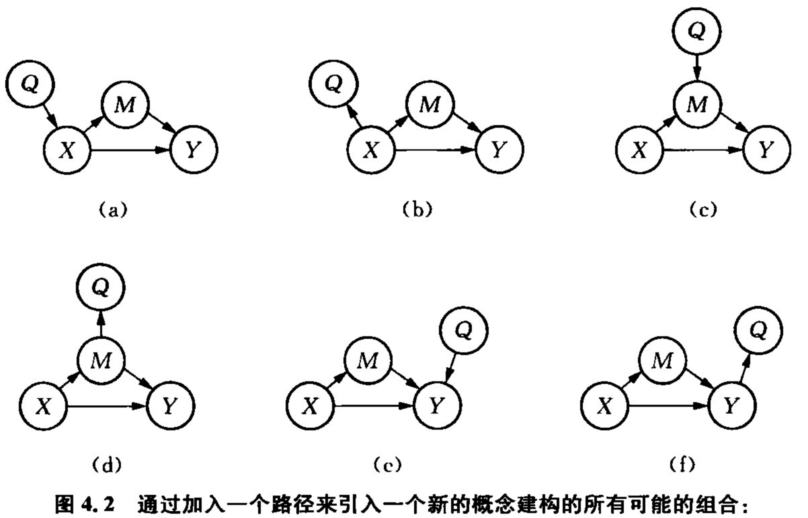



**图 4.2 通过加入一个路径来引入一个新的概念建构的所有可能的组合：$Q$ 作为 $X$、$M$ 和 $Y$ 的一个单独的原因或结果**

## 举例说明

让我们从前面一直用到的例子开始解释，在该例中，$r_{XM}=0.55$，$r_{XY}=0.45$，$r_{MY}=0.63$，得到的路径系数为 $a=0.55$，$b=0.55$，$c=0.15$。这些估计值显示在图 4.3 的右上角，代表了“基准模型”。现在让我们通过加入第四个概念建构 $Q$ 来增加这些相关关系。为了说明的简便，令 $r_{QX}=r_{QM}=r_{QY}=0.40$。在模型 a、模型 b、模型 d 和模型 f 中，3 个关键路径系数的估计值没有发生变化，而且，凡是涉及 $Q$ 的连接，其相关系数一律为 0.40，准确地反映了输入的信息。

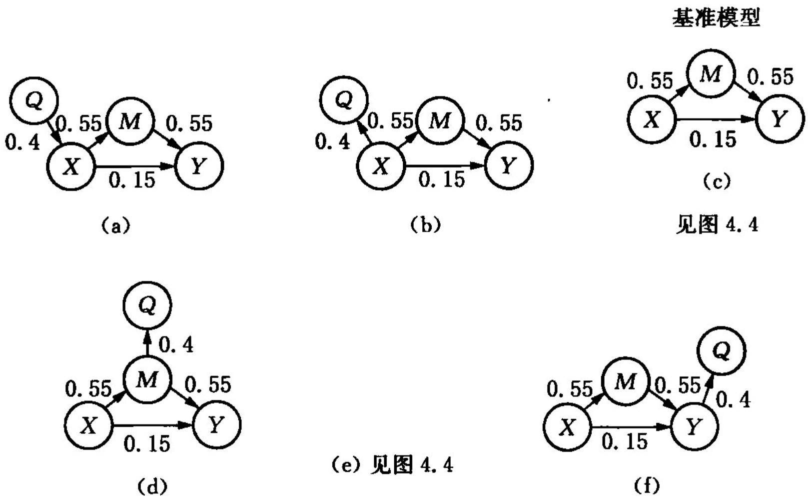



**图 4.3 $Q$ 对 $X$、$M$ 和 $Y$ 之间的中介作用的影响**

模型 c 和模型 e 见图 4.4。这两个模型的表现有所不同，因为将 $Q$ 作为 $M$ 或 $Y$ 的一个原因加入模型，意味着出现了两个外生的概念建构（$X$ 和 $Q$），这就带来了一个统计上的（也是概念上和经验上的）要求，即它们之间的相关关系也应该被呈现出来并加以估计。因此，首先要弄清楚的是，虽然（图 4.3 中的）模型 a、模型 b、模型 d 和模型 f 带来了两个自由度，但（图 4.4 中的）模型 c 和模型 e 则只多带来了一个自由度（另外一个自由度在估计两个外生概念建构之间的相关关系时被使用了）。

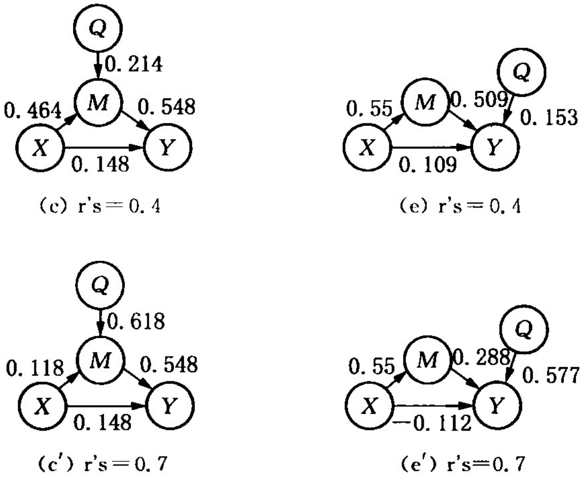



**图 4.4 不要将 $Q$ 作为 $M$ 或 $Y$ 的一个原因引入模型**

更成问题的是，3 个关键的中介作用路径系数不再是不变的了。在模型 c 中，$Q$（与 $X$ 一起）构成了 $M$ 的一个预测因素，结果是，$Q$ 和 $X$ 之间的多元共线性使得两个估计值分担了预测的方差，因此，$Q \rightarrow M$ 的路径系数不再与输入的信息 0.40 相同，而且 $X \rightarrow M$ 的路径系数也不再是我们所预期的 0.55（即不再代表研究总体的已知结构）。但要注意的是，与预测 $Y$ 有关的路径并没有受到影响。相反，在模型 e 中，由于 $Q$（与 $X$ 和 $M$ 一起）构成了 $Y$ 的一个预测因素，所以 $X$、$M$ 和 $Y$ 之间的多元共线性对涉及 $Y$ 的每一个路径都产生了影响，这些路径分别是：$Q \rightarrow Y$（我们对这一路径并不关心）、$X \rightarrow Y$ 和 $M \rightarrow Y$（我们关心的是这两个路径）。

为了便于说明，我们将所有的 0.4（即 $Q$ 与 $X$、$M$ 和 $Y$ 之间的相关关系）都替换为 0.7，模型 c′和模型 e′显示了这会对结果造成什么影响。多元共线性的增强产生了更加不同的结果（包括使其中一个作用路径的方向发生倒转，这表明，系统为了抵消新出现的高度线性相关而产生了一个抑制性的关系）。

从某种意义上来说，这些模型的结果肯定是“真的”，也就是说，当 $Q$ 与 $X$、$M$ 和 $Y$ 在任何程度上相关时（即只要相关度大于 0），路径系数的估计值就是上面所报告的那些数字。但是，如果研究者的视野主要关注的是 $X$、$M$ 和 $Y$，为了不扰乱最核心的关系，最好还是为 $Q$ 选定一个角色，也就是说，将 $Q$ 作为 $X$ 的一个原因引入模型，或将 $Q$ 作为 $X$、$M$ 或 $Y$ 的一个结果引入模型。[22]

上述结果概括了在哪些情况下 $r_{XM}$、$r_{XY}$、$r_{MY}$ 这 3 个核心指标会发生变化，而不管数据中的中介作用强度有多大（例如，是 25%还是 50%，或者是 75%）。而且，加入的相关关系 $r_{QM}$、$r_{QX}$、$r_{QY}$ 的强弱变化也会对结果造成影响（正如 0.40 与 0.70 的结果对比所显示的那样）。

为了理解这到底是怎么回事，一个简单的方法是将中介作用的实质理解为 $X$ 与 $Y$ 之间的偏相关在统计上控制了它们与 $M$ 的关系之后会变为 0，这一偏相关的计算公式为：

$$r_{XY\cdot M}=\frac{r_{XY}-r_{XM}r_{YM}}{\sqrt{(1-r_{XM}^2)(1-r_{YM}^2)}}$$

用文字表达就是："完全的中介模型是这样预测的，即 $X$ 对 $M$ 有一个直接的作用，$M$ 对 $Y$ 有一个直接的作用，并且当 $M$ 保持不变时，$X$ 与 $Y$ 之间不存在直接相关。"请回忆一下我们之前给出的提醒，这个偏相关在几种情况下是一致的：假设的 $X\rightarrow M\rightarrow Y$ 这一中介作用，倒转的因果关系链 $Y\rightarrow M\rightarrow X$，或者 $M$ 是 $X$ 和 $Y$ 的一个共同决定因素，即 $X\leftarrow M\rightarrow Y$（McDonald, 2001）。在 SEM 中，这些关系的等价性会导致在没有 $Q$ 的情况下，这些模型的总的拟合统计量将是相同的：所有的拟合优度测量（例如，$R^2$）都将等于 1，所有的拟合劣度都将等于 0。[23]然而，首先应该拟合我们所期望的中介模型，如图 1.1 所示，然后才继续拟合替代性的或竞争性的模型——这可以从 $Y\rightarrow M\rightarrow X$ 开始，再对起中介作用的概念建构赋予其他的角色，例如 $M\rightarrow X\rightarrow Y$ 或 $X\rightarrow Y\rightarrow M$——并且表明，虽然总的拟合统计量可能是相同的，但某些模型的参数估计却并不显著，或者不具有理论上的意义。

加入 $Q$ 之后，多余的自由度允许我们比较模型的拟合性。但是坦白地讲，在只多出两个自由度的情况下，会出现两个问题：首先，模型的拟合性可能只在某种程度上可以进行比较；其次，对太多的竞争性模型进行比较会导致过多的 Ⅰ类错误（Type Ⅰ errors）。若在经典的中介模型 $X\rightarrow M\rightarrow Y$ 中加入 $Q\rightarrow X$，则基本的拟合性指标如下：GFI = 0.96，RMR = 0.09，CFI = 0.96。当整个因果关系的方向发生倒转，即 $Y\rightarrow M\rightarrow X$（加入 $Q\rightarrow X$）时，模型的拟合性如下：GFI = 0.91，RMR = 0.20，CFI = 0.86。当检验 $M\rightarrow X\rightarrow Y$ 这一作用模式时，拟合指标

如下：GFI = 0.98，RMR = 0.038，CFI = 0.99。这一组指标表明，最后一个模型具有少许优势，尽管这正是我们所预料的。

如果没有额外的概念建构，最核心的三概念建构模型是"恰好识别"的，换一种说法是，许多竞争性的模型同样会很好地拟合数据。MacCallum 等人(1993)详细讨论了 SEM 方法中的模型识别和替代性模型的问题。即使统计上不存在等价模型这一问题，对希望取得进展的研究者来说，强有力的理论对于他们支持或反对某些形式的中介作用也是必不可少的。读者可以思考下面这段文字：

> 一个拟合性好的模型必须能够准确地再现观察到的数据。然而，即使发现了某一模型与数据具有一致性，几乎总有替代性的模型同样具备这种一致性……Stelzl(1986)发展了几条规则来帮助生成一系列等价模型……两个等价模型将会同样有效地解释同一个观察到的协方差模式，它们中的每一个，其总的拟合优度统计量将会相同……然而，两个等价模型对单个参数的估计值会有所不同，这就为两者的区分提供了一些基础。(Breckler，1990:262)

注意避免过度地将相关数据解释为中介作用(或其他的因果关系结构)的证据是相当重要的，用理论来补充经验结果也是重要的。有鉴于此，我们再提供两个例子来加以说明。首先，想象一下两种完全的中介作用的情形。其中一种情形是中介因素在概念上或是在时间上更接近于 $X$，这或许是因为对 $M$ 和 $X$ 的测量在时间上或在程序上如此接近，以至于方法变异量可能进一步促进了它们之间的联系。这种接近性或许可以被描述为 $X \rightarrow M \longrightarrow Y$，在经验上或许可以用如下值来表示：$r_{XM} = 0.90$，$r_{XY} = 0.30$，$r_{MY} = 0.30$。作为对比，另一种情形是中介因素更多地与 $Y$ 具有高度相关，这可以被描述为 $X \longrightarrow M \rightarrow Y$，假设具体的值是 $r_{XM} = 0.30$，$r_{XY} = 0.30$，$r_{MY} = 0.90$。如果我们对两组数据分别拟合中介模型，则第一种情形下的路径系数估计值为 $r_{MX} = 0.90(p < 0.05)$，$r_{YX} = 0.15(p > 0.05)$，$\beta_{YM} = 0.15(p > 0.05)$ (中介作用的比例为 47.4%)；第二种情形下的估计值为 $r_{MX} = 0.30(p < 0.05)$，$r_{YX} = 0.03(p > 0.05)$，$\beta_{YM} = 0.89(p <$

0.05)(中介作用的比例为89.9%)。这些参数估计值的显著与不显著的情况表明,中介作用在第二种情形下得到了支持,但在第一种情形下没有得到支持,尽管在概念上,两种情形都显示出存在中介作用的关系链。[24]

## 抑制作用

上面的第二种情形涉及抑制作用的问题,这在中介作用文献中已有所讨论。MacKinnon、Krull、Lockwood(2000)比较了中介作用和抑制作用:

> 在中介作用中,典型的情况是假设……自变量与因变量之间的关系……会减弱,因为中介变量解释了这一关系的部分或者全部……然而,可能出现的情形是,移除某一个变量可能会增加自变量和因变量之间的关系强度。这种改变就意味着存在抑制作用。(2000:174)

他们给出了一个例子:想象 $X$ 是对员工智力的测量,$M$ 是对他们感觉无聊的程度测量,$Y$ 是一条装配线上出现工作失误的次数。一位人力资源研究者会期望 $r_{XM} > 0$, $r_{MY} > 0$, 但 $r_{XY}$ 可能会是负的。

对中介关系(以及 SEM)框架中的抑制作用的一种诊断标准是,$a \times b$(即 $\gamma_{MX} \times \beta_{YM}$)的符号与 c(即 $\gamma_{YX}$)的符号相反(MacKinnon et al., 2000:174; Shrout & Bolger, 2002:432; Tzelgov & Henik, 1991)。许多社会科学家都听说过抑制作用,但对它并不是很熟悉,有鉴于此,另外一个例子在这里可能有用。这个例子是由 Shrout 和 Bolger(2002:431) 提供的。假定 $X$ 是某种压力源,$M$ 是一种应对行为,$Y$ 是所经历的焦虑程度。一名临床医生会期望 $r_{XM} > 0$, $r_{MY} < 0$, 于是中介的、间接的路径系数应该为负。最终,$r_{XY}$ 会是正的。因此,$a \times b$(表示中介路径)的符号与 c(表示直接路径)的符号会相反,这表明存在抑制作用。

另一种思考抑制作用的方法是将其与多元共线性进行对比。我们可以看到的多元共线性的例子通常是这样一种情况:“一个预测变量本来对结果变量具有较强的影响,但在控制了另一个预测变量之后,它的影响就变得微不足道

了。"而抑制作用则是"一种相反的情形，即一个预测变量的影响本来微不足道，但在控制了另一个预测变量之后，它的影响变强了"(Lynn，2003:58)。例如，令 $Y$ 为因变量，$A$ 为一个预测变量，$B$ 为另一个预测变量，并且，$B$ 与 $A$ 相关 ($r_{AB} \neq 0$) 但与 $Y$ 不相关 ($r_{BY} = 0$)。如果 $R^2$(即同时用 $A$ 和 $B$ 来预测 $Y$ 的多元回归方程的 $R^2$)大于两个简单相关系数的平方和，即 $R^2 > (r_{AY}^2 + r_{BY}^2)$，那么 $B$ 是一个起抑制作用的变量。

举这些例子的目的是要提醒可能会用到中介模型的研究者那句永恒的谚语"若输入错误的数据，则会输出错误的结果"在这里是同样适用的。若没有强有力的、清晰的理论总结，包括对竞争性理论进行检验的意图，中介分析的统计方法本身是不会变魔术般地澄清数据结构的。

在下面几节中，我们将用 LISREL 语句来结束本章内容，这些语句适用于我们所展示的超过 3 个关键概念建构的情形。在下一章中，我们会转向更加复杂的数据分析情形。

## 四概念建构模型的 LISREL 语句

本节将对图 4.3 所建议的 4 个模型中的每一个提供 LISREL 语句。在第一批语句中，$Q$ 是 $X$ 的一个原因：

```
Title: Four Constructs, Q→X(Figure4.3a).
da ni = 4 no = 100 ma = cm
la
x m y q
cm sy
1.00
0.30 1.00
0.30 0.30 1.00
0.30 0.30 0.30 1.00
```

```
mo ny=3 ne=3 nx=1 nk=1 lx=id, fi td=ze, fi ly=id, fi te=ze, fi
be=fu, fr ga=fu, fr
pa ga
1
0
0
pa be
0 0 0
1 0 0
1 1 0
out me=ml rs ef
```

在第二批输入指令中，$Q$ 是 $X$ 的一个结果：

```
Title: Four Constructs, X→Q(Figure 4.3b).
da ni=4 no=100 ma=cm
la
x m y q
cm sy
1.00
0.30 1.00
0.30 0.30 1.00
0.30 0.30 0.30 1.00
mo ny=3 ne=3 nx=1 nk=1 lx=id, fi td=ze, fi ly=id, fi te=ze, fi
be=fu, fr ga=fu, fr
pa ga
1
0
```

```
0
pa be
0 0 0
1 0 0
1 1 0
out me = ml rs ef
```

在第三个模型中,$Q$ 是 $M$ 的一个结果:

```
Title: Four Constructs, M→Q(Figure 4.3d).
da ni = 4 no = 100 ma = cm
la
x m y q
cm sy
1.00
0.30 1.00
0.30 0.30 1.00
0.30 0.30 0.30 1.00
se
m y q x
mo ny = 3 ne = 3 nx = 1 nk = 1 lx = id, fi td = ze, fi ly = id, fi te = ze, fi
be = fu, fr ga = fu, fr
pa ga
1
1
0
pa be
0 0 0
```

```
1 0 0
1 0 0
out me = ml rs ef
```

在第四个即最后一个模型中,$Q$ 是 $Y$ 的一个结果:

```
Title: Four Constructs, Y→Q(Figure4.3f).
da ni = 4 no = 100 ma = cm
la
x m y q
cm sy
1.00
0.30 1.00
0.30 0.30 1.00
0.30 0.30 0.30 1.00
se
m y q x
mo ny = 3 ne = 3 nx = 1 nk = 1 lx = id, fi td = ze, fi ly = id, fi te = ze, fi
be = fu, fr ga = fu, fr
pa ga
1
1
0
pa be
0 0 0
1 0 0
0 1 0
out me = ml rs ef
```

## 较长的中介作用链与替代性中介路径的 LISREL 语句

当第四个概念建构拉长了中介作用链时，即 $X \rightarrow M_1 \rightarrow M_2 \rightarrow Y$，就可以使用如下语句：

```
Title: X to M1 to M2 to Y.
da ni = 4 no = 100 ma = cm
la
x m1 m2 y
cm sy
1.00
0.30 1.00
0.30 0.30 1.00
0.30 0.30 0.30 1.00
se
m1 m2 y x
mo ny = 3 ne = 3 nx = 1 nk = 1 lx = id, fi td = ze, fi ly = id, fi te = ze, fi
be = fu, fr ga = fu, fr
pa ga
1
0
1
pa be
0 0 0
1 0 0
0 1 0
out me = ml rs ef
```

下面的语句检验的是两个分支中介路径 $X \to M_1 \to Y$ 和 $X \to M_2 \to Y$：

```
Title: X through M1 to Y and X through M2 to Y
da ni = 4 no = 100 ma = cm
la
x m1 m2 y
cm sy
1.00
0.30 1.00
0.30 0.30 1.00
0.30 0.30 0.30 1.00
se
m1 m2 y x
mo ny = 3 ne = 3 nx = 1 nk = 1 lx = id, fi td = ze, fi ly = id, fi te = ze, fi
be = fu, fr ga = fu, fr
pa ga
1
1
1
pa be
0 0 0
0 0 0
1 1 0
out me = ml rs ef
```

下面的 LISREL 语句显示了如何在一个更大、更复杂的分析网中，对中介作用进行检验，它具体对应的是图 4.1 中的结构：

```
Title: Nine Constructs, per Figure 4.1.
da ni=9 no=100 ma=cm
la
x m y a b c d e f
cm sy
1.00
0.30 1.00
0.30 0.30 1.00
0.30 0.30 0.30 1.00
0.30 0.30 0.30 0.30 1.00
0.30 0.30 0.30 0.30 0.30 1.00
0.30 0.30 0.30 0.30 0.30 0.30 1.00
0.30 0.30 0.30 0.30 0.30 0.30 0.30 1.00
0.30 0.30 0.30 0.30 0.30 0.30 0.30 0.30 1.00
se
x m y c e a b d f
mo ny=5 ne=5 nx=4 nk=4 lx=id, fi td=ze, fi ly=id, fi ly=id, fi
te=ze, fi be=fu, fr ga=fu, fr
pa ga
1 1 0 0
0 0 1 0
0 0 0 1
0 0 0 0
0 0 0 0
pa be
0 0 0 0 0
1 0 0 0 0
1 1 0 0 0
```

```
0 1 0 0 0
0 0 1 0 0
out me = ml rs ef
```

## 间接效应的显著性

中介作用最后要考虑的一个问题是区分 $X$ 对 $Y$ 的作用有多少是直接的（$X \to Y$），又有多少是通过中介因素而间接实现的（$X \to M \to Y$）。结构方程模型的使用者对这种分解的估计和显著性检验进行了探索（参见 Bollen，1987；Holbert & Stephenson，2003；Preacher & Hayes，2004）。

在简单的中介作用形式中（例如由图 1.1 首次展示的那样），$X$ 对 $Y$ 的直接效应是由系数 $c$（或者是 LISREL 模型中的 $\gamma_{YX}$）来估计的，$X$ 通过 $M$ 对 $Y$ 的间接效是由相应的结构系数的乘积（即 $a \times b$，或者是 LISREL 模型中的 $\gamma_{MX} \times \beta_{YM}$）来估计的，$X$ 通过直接和间接路径对 $Y$ 的总效应是两者之和：$[c+(a \times b)]$。

将总效应分解为直接效应和间接效应，这在大多数 SEM 软件中都可以通过一些选项或设定来实现。例如，在 LISREL 中，若在输出行中加入关键词“ef”，就可以得到直接效应、间接效应和总效应的估计值。

从统计计算软件包的输出结果可以得到直接效应、间接效应以及总效应，也可以得到这些单个项的显著性检验。理论上，方程 2.4 中的 $z$ 检验类似于间接效应的显著性检验。具体来说，这两个结果在本质上应该趋于接近（显著或不显著），即使实际数值存在某种程度的差别。最后，不管直接或间接效应显著与否，都有待研究者作出一个科学的判断，并对直接和间接作用路径的相对大小进行总结。

## 其他议题

如果研究者有额外的概念建构，如图 4.1 或图 4.3，就有更多的自由度可以使用。而当存在更多的自由度时，就可以获得比较性的拟合统计量，也就是两

个嵌套模型的 $X^2$ 值之差(即 $\Delta X^2$,类似于回归中的 $\Delta R^2$,例如,用 $X$、$M$ 来预测 $Y$ 的模型 1 嵌套于用 $X$、$M$、$Q$、$W$ 来预测 $Y$ 的模型 2,于是,两个模型的 $X^2$ 值之差 $\Delta X^2 = X_1^2 - X_2^2$,自由度之差 $\Delta df = df_1 - df_2$)。

层级嵌套模型并不是必须的,然而,拟合这些模型却为研究者提供了又一个优势。但问题在于,不管是用旧式的回归方法还是用更适当的结构方程模型来拟合,对中介作用的考察几乎肯定会在预测变量之间具有多元共线性的数据中进行。要作出中介作用存在的结论,其中一个条件是 $X$ 与 $M$ 有关,即 $X \rightarrow M$(中介作用的前半个路径)。于是,当关注 $Y$ 的预测和解释时,对直接路径(包含 $X$)和间接路径(包含 $X$ 和 $M$)的估计是通过两个本来就相关的预测变量来进行的,而这正符合多元共线性的定义。

多元共线性影响了结构方程模型中的结构路径系数 $\beta$ 和 $\gamma$(或者是回归中的 $\beta$)的估计值、显著性检验以及对它们的解释,但并不影响总的拟合性指标($X^2$ 或 $R^2$);多元共线性影响了将方差归因于不同的预测因素这一分配过程,但并不影响预测因素一共解释了多少方差这一总的评估。因此,对多元共线性有所担心的研究者可能不希望去研究结构路径参数的估计,而是研究嵌套模型的总的拟合性以及对它们进行比较的指标。遗憾的是,拟合嵌套模型的想法使路径建模方法回到了最初的用回归建模方法来拟合多个模型这一更加复杂的操作方式,而对这种方式的批评是,它导致了将苹果模型与柑橘模型进行比较和检验这种不恰当的情形。

对多元共线性的考虑和处理近似于另一种哲学上的观点,即由于人们感兴趣的任何数据都会自然而然地显示出某种多元共线性,因此需要对 $\beta$ 和 $\gamma$ 的估计值进行检验。拟合一个联立模型——它同时包括了直接联接($X \rightarrow Y$)和间接连接($X \rightarrow M \rightarrow Y$)以及其他任何概念建构之间的任意连接——可能确实包含了多元共线性,但同步拟合这一思想的精髓正是,所有的影响效应都是排除了或在统计上控制了模型中的其他效应后估计得到的。

# 第 5 章 | 高级议题

本章将考察一系列更加复杂的数据情形，研究者试图理解其中的中介作用关系。首先，我们将 SEM 应用于“中介性的调节作用”。一般说来，“调节作用”可以简单地解释为有一个交互项捕捉了一种条件性的关系。在中介性的调节作用中，交互项与因变量之间的关系是中介性的。

其次，我们将转向“调节性的中介作用”。在这种情形下，我们关心的问题是某些变量是否对中介作用具有调节性，或是否调和了中介作用的强度。最简单、最常见的情况是，对于两个群组来说，研究问题是要弄清楚中介作用是否只在一个群组中存在而在另一个群组中不存在。也就是说，如果在一个群组中有中介作用 $X \rightarrow M \rightarrow Y$，但在另一个群组中只有直接关系 $X \rightarrow Y$，那么这个分组变量就是调节项。调节项可能会更加复杂(例如，可以是一个连续变量，而不只是对两个群组进行区分)，我们对此也会加以考察。

第三，我们将考察研究者在遇到定类变量时应该怎么办。

第四，我们将探讨纵向数据所产生的问题。

## 中介性的调节作用

探究中介性的调节作用，就是考察涉及一个交互项的中介模型(Muller, Judd & Yzerbyt, 2005)。当进行一项实验时，研究者操控 $X_1$ 和 $X_2$ 两个因素，经典的方差分析会将因变量 $Y$ 模型化为由两个主效应以及它们的交互项 $X_1 * X_2$ 所构成的函数。同样，在中介分析中，调节作用也是由一个交互项来表示的，没有什么特别之处。

此时的中介作用分析与通常的情形一样，只不过涉及了一个调节项，也就是说，研究问题仍然相似，即到底存在一个直接作用$\{X_1 * X_2\} \rightarrow Y$还是一个间接的中介作用$\{X_1 * X_2\} \rightarrow M \rightarrow Y$。

如图 5.1 所示，主效应也被纳入模型，这是为了达到统计控制的目的（以排除低阶项以及与其有关的但却属于外来的方差和协方差），然而，它们并不是研究的关键和实质性兴趣所在。为了回答调节项是否具有中介关系这一问题，我们关注的是评估间接路径参数估计——$a \times b$（即$\gamma_{M, X_1 * X_2} \times \beta_{YM}$）——与直接路径参数估计$c$（即$\gamma_{Y, X_1 * X_2}$）的相对大小。[25]

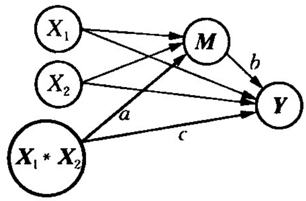



**图 5.1　中介性的调节作用**

如果理论要求检验中介性的调节作用，那么就应该用这些方法作为检验的手段。然而，虽然交互项在对实验中操控因素的方差分析中十分常见，同时也确实是研究者所期望的，但已经有文献表明，当概念建构被操作化为连续变量进而用它们的乘积项来表示其交互作用时，是存在一定问题的。应用包含连续型预测变量以及它们的乘积项的广义线性模型时所产生的问题，在这里也适用。研究者在生成乘积项之前应该特别对主效应变量进行均值对中处理，以减少乘积项与其组成部分之间的多元共线性，并且，研究者或许应该根据数据作图，以检验生成的乘积项在多大程度上导致了棘手的非线性特征，从而使线性模型可能不再稳健（James & Brett，1984：310），如此等等。[26]

如果预测变量应该由多项指标来表示，那么它们各自量表的均值就应该相乘以得到一个单一项来表示交互项。例如，$X_{1a}$、$X_{1b}$、$X_{1c}$会被简化为$\bar{X}_1$，$X_{2a}$、$X_{2b}$、$X_{2c}$被简化为$\bar{X}_2$，由此得到一个乘积项$\bar{X}_1 \times \bar{X}_2$。

总而言之，对中介性的调节作用来说，主效应因素基本上是被当做控制变量（这与方差分析相似）纳入模型的，因为研究者的兴趣更多集中在$X_1 * X_2$如何起

作用之上。至于 $X_1 \rightarrow M$、$X_1 \rightarrow Y$、$X_2 \rightarrow M$、$X_2 \rightarrow Y$ 这些路径是否显著则无关紧要。研究将这些估计置于一边，只关注 $X_1 * X_2 \rightarrow M$，$X_1 * X_2 \rightarrow Y$ 这一中介作用估计。

用来拟合图 5.1 所示模型的 LISREL 语句如下：

```
Title: X1, X2, and X1 * X2 each to M and to Y.
da ni = 5 no = 100 ma = cm
la
x1 x2 x12 m y
cm sy
1.00
0.30 1.00
0.30 0.30 1.00
0.30 0.30 0.30 1.00
0.30 0.30 0.30 0.30 1.00
se
m y x1 x2 x12
mo ny = 2 ne = 2 nx = 3 nk = 3 lx = id, fi td = ze, fi ly = id, fi te = ze, fi
be = fu, fr ga = fu, fr
pa ga
1 1 1
1 1 1
pa be
0 0
1 0
out me = ml rs ef
```

最后请注意，我们有足够的自由度来估计这个模型。模型中一共有 5 个概念建构，所以，我们会从 10 个协方差中提取一部分来估计 7 个路径。

## 调节性的中介作用

“调节性的中介作用”这一概念反映的是这样一种想法，即中介关系是以某一调节变量的取值为条件的。例如，中介作用只存在于一个样本而非另一个样本，或者中介作用有可能随着其他某一变量的增加而增强等等(Muller et al.，2005)。

典型的情况是，调节变量是一个定类变量；更常见的情况是，该定类变量只取两个值。也就是说，有两个群组，此时研究者的兴趣是要确立证据表明在一个样本组中存在某种中介关系，而在另一个样本组中则存在某种直接关系。对群组的定义可以依据实验所设定的条件，也可以依据个体的区别性变量，例如性别或者以特征测量量表的中位值对人群进行划分等等(调节变量为连续型变量的情况较为少见，其处理方法会在本章的末尾加以讨论)。

用旧式的回归方法来检验调节性的中介作用是相当困难的。回归分析技术会用到定义组别的虚拟变量以及它们与各路径系数的交互项，但这种方法相当笨拙。SEM 自然而然地有一个在方法上与此对应的技术可以实现上述实质性的分析任务。在 LISREL 和其他 SEM 软件中，都有语句选项可以用来拟合多组结构方程模型。以图 5.2 为例，该模型说明的是，每个群组都有 3 个路径，但本质上的理论预期是，直接连接 $c$ 在一个群组中显著，而由 $a$ 和 $b$ 的估计值所表示的间接路径却在另一个群组中显著。

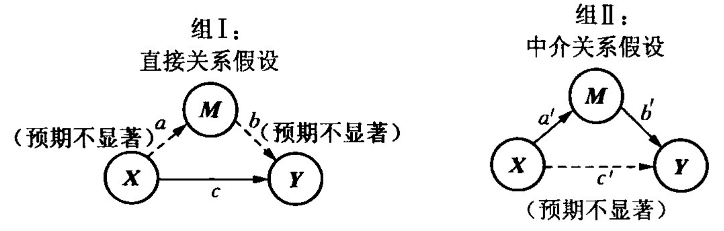



**图 5.2 调节性的中介作用**

对该模型的估计过程是，输入第一个协方差矩阵，用所有 3 个(直接和间接)路径来设定模型。接着在语句中(下文很快就会给出)输入第二个协方差矩阵，

这时使用者既可以将两个群组中的系数模式设定为相同，也可以将跨组系数设定为相等（即它们跨组“不变”）。

如果研究者只有 3 个关键概念建构 $X$、$M$ 和 $Y$，要求在两个样本组中都拟合同一个关系模式会导致两个样本组都出现完全拟合（原因在于前面所说的模型识别的问题）。参数估计值可能会有所差异，并且这些差异可能足以使研究者对两个群组作出区分，即在一个群组中确立了中介关系而在另一个群组中确立了直接连接。例如，我们用 $p_1$ 表示组Ⅰ中由中介作用所解释的 $Y$ 的变异比例，$p_1=(a_1\times b_1/(a_1\times b_1)+c_1)$，用 $p_2$ 表示组Ⅱ中由中介作用所解释的 $Y$ 的变异比例，$p_2=(a_2\times b_2/(a_2\times b_2)+c_2)$，我们可以用一个标准的 z 检验来比较这两个比例的相对大小：

$$z=\frac{p_1-p_2}{\sqrt{\frac{p_1(1-p_1)}{n_1}+\frac{p_2(1-p_2)}{n_2}}}$$

若 $z>1.96$，则组Ⅰ中的中介作用比例较大；若 $z<-1.96$，则组Ⅱ中的中介作用比例较大。若 z 不显著，则两组中的中介路径强度相当。

在只有 $X$、$M$ 和 $Y$ 3 个概念建构的情况下，另一种替代性的方法是试图（为两个群组）寻找一个（共同的）拟合统计量的研究者必须对其不变性进行检验。在这里，参数被设定为相等，$a=a'$，$b=b'$，$c=c'$，并且，试图展示调节性的中介作用的研究者会期望统计量表明该模型（即将参数设定为相等的模型）并不拟合数据。

当研究者面临的情形是 $X$、$M$ 和 $Y$ 嵌入一个更复杂的分析网（即加入 $Q$）时，“不变性模型”会被估计，并且拟合统计量应该又一次表明该模型无法拟合数据。假如确实如此，使用者这一次就应该重新运行 LISREL，要求估计同样的“模式”以得到明显不同的参数估计值，并查看哪一组展示出了更强的中介效应。

在两个群组中拟合基本的三变量中介模型的 LISREL 语句如下：

```
Title: Moderated Mediation.
da ng = 2 ni = 3 no = 100 ma = cm
```

```
la
x m y
cm sy
1.00
0.30 1.00
0.30 0.30 1.00
se
m y x
mo ny=2 ne=2 nx=1 nk=1 lx=id, fi td=ze, fi ly=id, fi te=ze, fi
be=fu, fr ga=fu, fr
pa ga
1
1
pa be
0 0
1 0
out me=ml rs ef
da ni=3 no=100 ma=cm
la
x m y
cm sy
1.00
0.30 1.00
0.30 0.30 1.00
se
m y x
mo be=ps ga=ps
out me=ml rs ef
```

在上述语句中，第二个模型（"mo"）陈述句使用的是"ps"或"same pattern"选项（为了在两个群组中得到很可能不同的参数估计值）。当研究者拟合"不变性模型"时，则应该使用"in"选项。

让我们来看一个调节性中介作用的例子。建立一组数据，使其在组Ⅰ中显示出75%的中介作用，在组Ⅱ中显示出25%的中介作用，上述模型恰好能够拟合此数据（$X_3^2 = 7.05$，$p = 0.07$，CFI = 0.97）。所以，我们可以得出以下结论：不管我们看到的数据形式是什么，两个群组中的中介作用大小在本质上具有统计等价性。一个具有比较性的结果是，当我们检验中介作用在组Ⅰ中为100%而在组Ⅱ中为0%这一模型时，它显然无法拟合数据（$X_3^2 = 22.01$，$p = 0.00$；CFI = 0.57）。我们知道，该模型所表示的情形是中介作用的强度在两个群组中有所不同，因此，它的欠拟合性就表明，组Ⅰ中的关系结构与组Ⅱ中的关系结构显著不同。

对调节性的中介作用进行这种处理的前提是调节变量为定类变量，不管该定类变量是取两个值还是取更多的值，多组分析都能够轻易地容纳两个或多个群组。然而，当调节变量并非一个离散变量时，就产生了如何进行建模的问题。下面我们就来解答这一问题。

假设我们现在只需对 $X \to Y$ 进行建模。如果我们提出这样一个问题："这一关系是调节性的吗？"我们就会引入一个调节变量的主效应（这是出于统计控制的原因）和交互效应，于是一共引入了两个新路径：$Mod \to Y$ 和 $\{X * Mod\} \to Y$。

在完整的中介作用情形下，我们的处理过程与前面类似。即，如果调节性的中介作用中的调节变量是连续型的，我们就引入 $Mod \to Y$ 和 $\{X * Mod\} \to Y$，$Mod \to M$ 和 $\{X * Mod\} \to M$ 以及 $\{M * Mod\} \to Y$。该模型如图 5.3 所示。

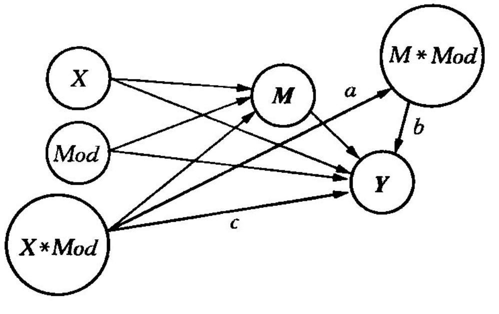



**图 5.3 调节性的中介作用：连续型调节变量的情形**

图 5.3 与图 5.1 有部分类似，但两者不应被混淆。在图 5.1 中(即中介性的调节作用)，所有的主效应和交互效应都被纳入模型，但中介分析只关注“通常的”3 个参数，即 $a$、$b$、$c$。在图 5.3 中，项目更多，并且它们几乎都只是被用来实现统计控制的。图中突出显示的路径 $a$、$b$、$c$ 是我们所关注的，但对它们的解释会稍微复杂一些。

假设上述模型中的调节变量是外向性的一个测量，较高的数值表示受访者更加外向。当路径 $X \rightarrow M$ 对个性外向的人来说更强时，系数 $a$ 会为正；当路径 $M \rightarrow Y$ 对个性外向的人来说更强时，系数 $b$ 也会为正；当路径 $X \rightarrow Y$ 对个性外向的人来说更强时，系数 $c$ 也会为正。路径系数 $a$、$b$、$c$ 的强度仍然可以用方程 2.4 中的 $z$ 检验来比较。

显然，这一分析技术包含了对一个只有两种取值的调节变量的建模，对后一种情况的分析应该(大体上)类似于前面讨论过的多组分析。假设根据外向性测量的中位值将样本一分为二(就使原来的连续型调节变量转化成了一个二元取值变量)，将内向的人编码为 0，将外向的人编码为 1。在这种情形下，如果对个性外向的人来说有路径 $X \rightarrow M$，那么系数 $a$ 就是正的；如果对个性外向的人来说有路径 $M \rightarrow Y$，那么系数 $b$ 也是正的；如果对个性外向的人来说有路径 $X \rightarrow Y$，那么系数 $c$ 也是正的。编码(0/1)规则可以反过来，如果这么做的话，这种再分析就等于是从个性内向的人的视角对数据进行了互补性的解读。

请注意，我们有足够的自由度来估计这一模型。模型中有 6 个概念建构，所以，15 个协方差为估计 8 个路径提供了足够的信息。

## 定类变量

用来检验中介结构的传统方法和 SEM 方法可以很好地处理连续型变量(即使是以五分或七分法测量的变量)。然而，社会科学中的许多重要变量都越来越多地在本质上是离散的或定类的(例如，是/否)。在本节中，我们会阐释如何在中介作用研究中处理定类变量。

例如，在行为决定理论(或者判断和决定过程)中，一名研究者可能对如下

问题感兴趣：一种人格特征——例如风险偏好($X$)——是否通过当前的生活评估($M$)影响了参与极限运动的决定($Y$)(这里$Y$是定类变量)？或者，专家意见($X$)是否通过品牌选择($M$)影响了顾客满意度($Y$)(这里$M$是定类变量)？同样，政治意识($X$)可能是连续型的，政党归属($M$)是离散型的，对总统某项行动的态度($Y$)是连续型的(Winship & Mare, 1983)。

有时，信息本来就是定类的，例如性别或种族。有时，连续型信息被进行归类处理，例如，将家庭收入转化为3个或4个阶级以反映社会经济地位，或者是根据测量受访者某些特征的量表得分中位值将样本一分为二等等。[27]

在中介分析中，如果$X$是定类变量，则问题不是很大——路径$X \rightarrow M$和$X \rightarrow Y$可以通过结构方程模型中的虚拟变量来进行估计。然而，当中介变量$M$或因变量$Y$是定类变量时，会带来较大的挑战。我们首先提出问题所在，然后提供解决方案。

## 定类变量的问题实质

我们首先考虑一个经典的情形，即将潜在的连续变量作为定类变量处理会导致"信息损失"，其结果是，涉及该定类变量的经验关系的强度被减弱了。表5.1包括了8个反映中介作用的结构方程模型的结果。我们生成了含有10000个观察个案的数据，并使其显示出25%的中介作用。[28]第一行，即模型A，将3个变量$X$、$M$和$Y$都处理成其原有的连续型形式，因此，其结果在本质上反映了数据的真实情况，而我们将会用其他的数据处理方式与其进行比较。

**表5.1　将连续变量$X$、$M$、$Y$处理为二元取值变量**

| 模型 | $X$ | $M$ | $Y$ | $a$ $X \rightarrow M$ | $c$ $X \rightarrow Y$ | $b$ $M \rightarrow Y$ | $ab/(ab+c)$ | $R^2_M$ | $R^2_Y$ |
|---|---|---|---|---|---|---|---|---|---|
| A | 连续 | 连续 | 连续 | 0.321 | 0.298 | 0.309 | 0.250 | 0.103 | 0.244 |
| B | 连续 | 连续 | 定类 | 0.321 | 0.248 | 0.249 | 0.244 | 0.103 | 0.163 |
| C | 连续 | 定类 | 连续 | 0.252 | 0.338 | 0.237 | 0.150 | 0.064 | 0.210 |
| D | 连续 | 定类 | 定类 | 0.252 | 0.276 | 0.206 | 0.158 | 0.064 | 0.148 |
| E | 定类 | 连续 | 连续 | 0.263 | 0.230 | 0.345 | 0.283 | 0.069 | 0.214 |
| F | 定类 | 连续 | 定类 | 0.263 | 0.192 | 0.278 | 0.276 | 0.069 | 0.142 |
| G | 定类 | 定类 | 连续 | 0.202 | 0.267 | 0.268 | 0.169 | 0.041 | 0.172 |
| H | 定类 | 定类 | 定类 | 0.202 | 0.218 | 0.232 | 0.177 | 0.041 | 0.122 |

在表格的其他部分，即模型 B 至模型 H，每一个变量都被注明在建模时是使用了原有的连续型形式，还是将其处理为定类变量，即用通常的中位值划分法将其转化为一个二元取值变量。对各个模型的比较只列出了值得注意的那些结果，但正如表格所示，这些比较结果并无令人惊异之处。

比较标记为“$a$”的那一列中的数字，它表示的是路径 $X \to M$ 的参数估计值。当 $Y$，即没有包括在该路径中的变量，被处理为定类变量时（即模型 B），路径 $X \to M$ 的系数并不受到影响（与模型 A 相比较）。但当 $M$ 被处理为定类变量时（即模型 C 和模型 D），或者当 $X$ 被处理为定类变量时（即模型 E 和模型 F），该估计值变小了。当 $M$ 和 $X$ 都被处理为二元取值变量时（即模型 G 和模型 H），与使用原有的连续型形式的模型相比，该估计值进一步变小。在对应于路径 $X \to Y$ 或 $M \to Y$ 的列中，有类似的观察（例如，对路径 $X \to Y$ 来说，将模型 A 分别与模型 B、模型 E 和模型 F 进行比较；对路径 $M \to Y$ 来说，将模型 A 分别与模型 B、模型 C 和模型 D 进行比较）。[29]

另一种思考“信息损失”的方式是一个二元取值变量（本来就是二元取值或者本来是连续的但后来被转换成二元取值）的方差变小了（例如，极端的情况是 $\sigma_X^2 = 0$）。因此，它作为预测变量来解释某一因变量变异的潜力就变小了。

表 5.1 的中介模型还可以有另外一种不同的观察方式。假如比较的关系多于一个，则当一个路径减弱时，总效应似乎会使另一条路径有所强化。例如，在标记为“$c$”的那一列中，当中介变量 $M$ 被处理为定类变量时，对路径 $X \to Y$ 的估计相对变强了（比较模型 A 和模型 C）。在 $M$ 上的信息损失削弱了 $M$ 与其他变量之间的关系（具体来说，是 $X \to M$ 和 $M \to Y$），但却有效地强化了除 $M$ 之外的变量之间的关系（特别是 $X \to Y$）。同样，在第三列中，当 $X$ 被处理为定类变量时，路径 $M \to Y$ 被强化了（比较模型 A 和模型 E）。这些结果虽然是“正确的”，但它们应该提醒了研究者，对一个变量的处理方式明显会影响到（数据中的）变化是直接从 $X$ 到 $Y$ 还是通过 $M$ 间接发生的这一结论。

假如一名研究者希望得出这样的结果，即中介路径是如此之弱，以至于应该将概念图示看做是直接的（即 $X \to Y$ 支配了 $X \to M \to Y$），那么他/她只需简单地将中介变量 $M$ 进行定类处理，就能为这种说法提供经验依据。相反，持中介

作用假设的研究者可以将 $X$ 处理为离散变量，就像有些研究者用实验中的操控因素或者将群组作为自变量一样。当然，对数据进行这种处理会掩盖潜在的真实效应并产生误导性的结果。因此，这反而提出了一种道德上的要求，即试图支持某种直接效应的研究者可能需要确定中介变量 $M$ 是连续的而不是离散的，而支持某种中介路径的研究者需要确定 $X$ 是连续变量。至少这些研究者应该设定更高、更保守的准则来确保中介作用统计量是清楚明确的。

上述阐释的目的并不是批评将连续变量处理为定类变量来使用的做法。有时候，研究者就只能获得分类信息；有时候，研究者在进行某项理论验证时，只需要在逻辑上对几组分类数据进行比较；有时候，连续型的测量看起来太过精确，因此反而具有误导性，如此等等。研究者必须使用他们已有的变量既有形式，或者是适合于他们的总体理论目标的变量形式。[30]上述阐释的目的是要再次提醒研究者反对这样一种倾向，即将相关性数据（以及由抽样、测量误差、样本截断等原因而造成的有误数据）过于急切地解释为一种因果关系链的证据。

## 解决定类变量的问题

在本节中，我们将探索 SEM 方法中与定类数据有关的一些技术思想。接下来会作具体阐述，但之后我们会用一些较简单的技术来进行总结，这会使那些不擅长技术的读者感到宽慰。

这类困境的一种解决方法是在中介分析中对 3 个变量进行类似的处理，即全部是连续变量或者全部是定类变量。这样一来，它们在测量方式上的不同就不会导致在对直接路径和中介路径的相对强度进行结论时出现虚假证据。我们已经知道如何通过 SEM 来处理 3 个连续变量。如果 3 个变量都是定类变量，那么对数线性模型就是合适的。logit 模型之所以是合适的，是因为对中介作用的探索是一项预测性的工作，即我们感兴趣的是对因变量的预测（Menard，1995）。但是，将 SEM 对联立方程进行建模的逻辑转移至 logit 或对数线性模型的分析情境并不简单。让我们来看一下究竟存在什么问题。

若要研究预测变量（如 $X$ 和 $M$）对一个二元取值因变量 $Y$ 的影响效应，则我

们会问："对于一个给定了 $X$ 和 $M$ 任意取值组合的观察个案来说，它的 $Y$ 值取 1 而不是 0 的可能性有多大（1 通常表示我们感兴趣的行为的出现）？"具体来说就是，当给定了 $X$ 和 $M$ 的特定取值后，$Y=1$ 相对于 $Y=0$ 的发生比是多少？对这些发生比以比率的形式进行比较，就得到了"比值比"，即 $P(Y=1)/(1-P(Y=1))$。相乘项总是比相加项更令人费解，因此，如果我们（对发生比）取自然对数，得到所谓的 logit：

$$\text{logit}(Y) = \ln(\text{odds}) = \ln\left[\frac{P(Y=1)}{1-P(Y=1)}\right] \qquad [5.1]$$

就可以将模型写为预测变量的一个函数，而这个函数具有大家所熟悉的线性形式：

$$\text{logit}(Y) = \beta_0 + \beta_X X + \beta_M M \qquad [5.2]$$

拟合这个 logit 模型会得到拟合统计量以及与路径 $X \rightarrow Y$ 和 $M \rightarrow Y$ 相应的参数估计。

$X$ 和 $M$ 之间的关系实际上也被包括在这个模型中，只是没有明确地写出来（正如方程 2.3 一样）。与中介分析的回归方法一样，另一个模型 $\text{logit}(M) = \beta_0 + \beta_X X$ 需要单独被拟合，以明确地分解出 $X \rightarrow M$ 这一路径。例如，Knoke 和 Burke（1980：43）描述了一个多方程方法用于对离散数据拟合因果关系模型，这种模型拟合是一波接一波地进行的，例如，在我们的中介分析应用中，先是 $X \rightarrow M$，再是 $M \rightarrow Y$。尽管这两个模型都是在相同的统计框架（即 logit）中拟合的，但其结果并没有被放置在一起，因为得出的"路径系数"无法被轻易地整合和解释（就像 SEM 中的 $a \times b$ 这一间接路径的情形）。Fienberg（1981）阐释了一个类似的例子，其中，成批的前兆变量被放入之后的分析阶段中，它们一起预测了终极内生变量。他对"路径系数"的解释持保守态度，建议研究者只报告其方向而不报告其估计值。[31]然而，如果我们倒转 logit 方程的原有形式为：

$$\text{logit}(Y) = \ln\left[\frac{P(Y=1)}{1-P(Y=1)} = \beta_X X + \beta_M M\right]$$

以求解如下公式：

$$P(Y=1) = \frac{e^{\beta_X X + \beta_M M}}{1 + e^{\beta_X X + \beta_M M}}$$

并假设路径 $X \to M$ 和 $M \to Y$ 相互独立以取它们各自概率的乘积(就像在 SEM 中计算 $a \times b$ 一样)——称其为 $p_m$(即中介作用的概率),我们就可以通过一个标准的 $z$ 检验来比较这一概率与数据沿直接路径 $p_d$ 变化的概率:

$$z = \frac{p_d - p_m}{\sqrt{\frac{p_d(1-p_d)}{n_d} + \frac{p_m(1-p_m)}{n_m}}}$$

虽然定类数据的因果关系模型可以按顺序依次拟合(就像旧式的回归方法),但这样做会使我们回到拟合多个模型的情形,而这种统计分析是次优的。关于这一点,我们在前面已经讨论过。

Collins、Graham 和 Flaherty(1998)对含有二元取值变量的数据采取了一个不同的方法来分析中介作用链。他们构建了一个从 $X$ 到 $M$ 到 $Y$ 的判定树,每一个点上所做的选择在概率上决定了它在哪一个点上终结。这种概念图示看似特别适合中介变量为二元取值的情况。如果我们将中介作用看做由下列连接构成:先是 $X \to M$(是)或 $X \to M$(否),然后是 $M$(是)$\to Y$(是)或 $M$(是)$\to Y$(否)以及 $M$(否)$\to Y$(是)或 $M$(否)$\to Y$(否)。由此我们就得到一个判定树,它从 $X$ 开始,首先在 $M$ 上分叉(是或否),然后再一次在 $Y$ 上分叉(是或否),如此则产生了 4 种终点状态:

1. $X \to M$(是) $\to Y$(是)
2. $X \to M$(是) $\to Y$(否)
3. $X \to M$(否) $\to Y$(是)
4. $X \to M$(否) $\to Y$(否)

[5.3]

如何拟合这 4 种状态是个问题,因为这不同于二项 logit 模型中只有两种状态的情况。然而,我们对这 4 种状态的兴趣并非完全相同。状态 1 代表了中介作用和间接路径;状态 3 表示的是从 $X$ 到 $Y$ 这一直接路径;状态 2 无关紧要(中介作用的其中一个路径得到支持,但另一个却并未得到支持);状态 4 也没有什么意义(直接和间接作用都不明显)。因此,虽然乍看之下 4 种状态会要求更复杂的建模,但其实只有两种状态有意义——状态 1 和状态 3。logit 模型可以轻

易地对这两种状态进行建模，将状态 2 和状态 4 排除在外（当成缺失数据丢弃）即可。logit 模型所得到的参数表示了间接路径相对于直接路径的“发生比”。一个显著为正的 $\beta_X$ 意味着中介作用得到支持；若估计值显著为负，则意味着直接路径得到支持；对一个不显著参数的一种解释则是存在部分中介作用。遗憾的是，对不显著结果总存在另一种解释，即缺少统计效力。

对所有 4 种状态的建模感兴趣的研究者可以运行一个多项 logit 模型，也就是针对含两个以上类别的因变量的 logit 模型。例如，想像一个因变量有 3 种选择（这可以简单地延伸至我们所说的 4 种状态）：$A$、$B$、$C$。多项 logit 模型用一组预测变量“$x$”来对成组的类别进行建模。为了在概念上得到简化，经常是对相邻的一组类别进行建模，它们之间的关系为（参见 Long，1997：150）：

$$\ln\left[\frac{P(A \mid x)}{P(B \mid x)}\right]+\ln\left[\frac{P(B \mid x)}{P(C \mid x)}\right]=\ln\left[\frac{P(A \mid x)}{P(C \mid x)}\right]$$

我们可以将状态 1 到状态 4 作为 $X$ 的一个函数来建模（$M$ 在创建 4 种状态时已经被包括在内了）。

不管是对我们讨论的还是其他的统计建模方法来说，中介作用的探究最终仍然是化简为直接路径和间接路径的强度比较。Winship 和 Mare（1983：83—84）根据 SEM 的方程（方程 3.20）$\hat{Y} = \Gamma X + BY + \Psi$，即

$$\begin{bmatrix}\hat{M}\\ \hat{Y}\end{bmatrix}=\begin{bmatrix}\gamma_{MX}\\ \gamma_{YX}\end{bmatrix}[X]+\begin{bmatrix}0 & 0\\ \beta_{YM} & 0\end{bmatrix}\begin{bmatrix}M\\ Y\end{bmatrix}+\begin{bmatrix}\Psi_M\\ \Psi_Y\end{bmatrix}=\begin{bmatrix}\gamma_{MX}X+\Psi_M\\ \gamma_{YX}X+\beta_{YM}M+\Psi_Y\end{bmatrix}$$

将 $X$ 对 $Y$ 的总影响效应分解为两个核心部分：$X$ 对 $Y$ 的直接作用与 $X$ 通过 $M$ 对 $Y$ 的间接作用。当 $Y$ 是连续变量时，我们知道这一总的影响效应是 $c+ab$，用 SEM 的标记方法就是 $\gamma_{YX}+(\gamma_{MX}\times\beta_{YM})$。相反，如果 $Y$ 是离散变量，则 Winship 和 Mare（1983：83—84）从下面的这一基本原则开始：

$$\frac{dY}{dX}=\frac{\partial Y}{\partial X}+\frac{\partial Y}{\partial M}\frac{dM}{dX}$$

在给定 $X$ 对 $Y=1$ 的概率的期望效应下（假设因变量只有两种取值）代入 logit：

$$\frac{dp(Y=1)}{dX}=\frac{\gamma_{YX}e(\gamma_{YX}X+\beta_{YM}M)}{[1+e(\gamma_{YX}X+\beta_{YM}M)]^2}+\frac{\beta_{YM}\gamma_{MX}e(\gamma_{YX}X+\beta_{YM}M)}{[1+e(\gamma_{YX}X+\beta_{YM}M)]^2}$$
$$=\gamma_{YX}p_Y(1-p_Y)+\beta_{YM}\gamma_{MX}p_Y(1-p_Y) \qquad [5.4]$$

其中，$p_Y=p(Y=1)$（当 $Y$ 本来是连续的但却用一个离散指标来测量时，情况会稍微复杂一些），并且需要在 $Y$ 的各个值上来衡量。然而，既然方程 5.4 右边的两部分都出现了 $p_Y(1-p_Y)$，那么探究的焦点仍然是对 $\gamma_{YX}$ 和 $\beta_{YM}\gamma_{MX}$ 相对大小的检验。

理解了这些问题之后，研究者又该怎么办呢？或许有一个简单的解决方案。对 3 个变量拟合一个包括了所有二维效应的对数线性模型，可以实现对路径 $X\rightarrow Y$ 和 $M\rightarrow Y$（正如方程 5.2 中的 logit 模型所做的那样）以及路径 $X\rightarrow M$ 的估计和检验。这一模型是这样的：

$$\ln(f_{xmy})=\beta_0+\beta_X+\beta_M+\beta_Y+\beta_{MX}+\beta_{YX}+\beta_{YM} \qquad [5.5]$$

该模型会生成拟合统计量，并给出 $\beta_{MX}$、$\beta_{YX}$、$\beta_{YM}$ 的估计值。[32]

针对表 5.1 中的定类数据，对数线性模型给出的估计值见表 5.2。[33]独立卡方具有叠加性，因此，中介连接相对于直接路径的大小可以用下面的方程来近似地检验：

$$(X^2_{MX}+X^2_{YM})-X^2_{YX}=\Delta X^2_{due\text{-}to\text{-}mediation}\sim\chi^2_1$$

在这组数据中，$\Delta X^2_{due\text{-}to\text{-}mediation}=(191.57+540.09)-482.98=248.68$，具有统计显著性，$p<0.05(\chi^2_{1;\,0.05}=3.841)$，这意味着部分中介作用得到支持（某些变化是直接的，因为 $X\rightarrow Y$ 这一连接显著）。

**表 5.2 表 5.1 中 X、M 和 Y 均为定类变量的对数线性参数估计**

| 模型参数 | 估计值 | $X^2$ |
|---|---|---|
| $\beta_{XM}$ | 0.296 | 191.57 |
| $\beta_{XY}$ | 0.471 | 482.98 |
| $\beta_{MY}$ | 0.493 | 540.09 |

注：每一个 $\chi^2$ 都显著，$p<0.0001$。

其他技术在文献中也有所探讨。其中,最具包容性的技术或许是使用了多分格相关的分析框架(适用于定序变量),由 Muthén(1984)提出。他的这一技术在方法上相当灵敏,有助于决定类别数量的阀值,但要求样本规模较大。其他技术发展的初步结果表明,一种可能有效的方法是拟合一个对数线性模型并倒转其估计参数的协方差矩阵(Kupek, 2005),再将其输入结构方程模型。

上述解决方案都是针对 3 个离散变量提出的。还存在的问题就是,当 $X$、$M$ 和 $Y$ 在测量形式上互不相同时(即其中的一个或两个变量是连续的,另外两个或另外一个变量是离散的),该如何解决? logistic 回归是 logit 模型的一个延伸,也是针对离散型因变量的,但却允许加入连续型或离散型的预测变量。因此,logistic 回归能够将离散型的 $M$ 或 $Y$ 建模为混合型预测变量的一个函数,但问题是,不同模型之间的整合很困难,即使都是 logistic 模型。更糟糕的是,如果一个内生变量是连续的(要求一个一般意义上的线性模型),而另一个是离散的(要求一个 logit 模型或 logistic 回归),那么它们之间是不可比的,因为对路径的估计会通过相当不同的方法来实现,包括使用不同前提的技术和以不同的测量尺度为基础来估计参数(即线性与发生比的对数比较)。这种问题甚至也不能用由不同阶段的方程所组成的混合模型来解决,例如层级嵌套模型,因为线性(或 logit)函数的输出结果会作为 logit(或线性)函数的输入条件。[34]

如前所述,对定类变量的主要考虑或担心是其固有的"信息损失"问题。这种信息损失已经被证明会导致对因子载荷(对定类变量而言)的低估和对卡方值的高估(DiStefano, 2002; Schumacker & Beyerlein, 2000)。请注意,这两种结果都在统计上犯了保守性的错误。一些模拟结果显示,基于定类变量与基于连续变量的结果具有相似性,除非是在二元取值变量的分布向相反的方向倾斜这种极端情形下(Ethington, 1987)。

使用定类数据的研究者的底线是,如果所有变量都是定类的,就可以按照方程 5.5 拟合一个对数线性模型。一般来说,如果 $X$、$M$ 和 $Y$ 是任意一种定类变量与连续变量的混合,则 SEM 看似为首选的解决方法(用虚拟变量来表示定类变量),正如我们之前所演示的那样。SEM 或许会损失某些关系,但它至少是一个保守的统计方法,它允许连续变量和离散变量的混合(在这一点上与 logis-

tic 回归相同，但它还可以进一步与其他模型进行比较，而不需要在 SEM 和 logit 模型之间来回转换）。

## 纵向数据

研究者有时希望用重复测量或同一组个体数据来估计中介模型。目前，已经有一些方法来应对这一新的考虑(参见 Cole & Maxwell, 2003; Judd, Kenny & McClelland, 2001)，但重复测量数据又可以被轻易地纳入结构方程模型。

图 5.4 显示了重复测量数据的实质——对 $X$、$M$ 和 $Y$ 的测量在时点 1 和时点 2 上分别获得，研究者据此探究通常的结构参数，即估计 $X_1 \rightarrow M_1 \rightarrow Y_1$、$X_1 \rightarrow Y_1$、$X_2 \rightarrow M_2 \rightarrow Y_2$、$X_2 \rightarrow Y_2$。新的复杂性在于，对数据收集于同一组个体这一本质性特征来说，需要处理误差的结构性相关。具体而言，测量误差被允许在 $X_1$ 和 $X_2$ 之间相关，即 $\theta_{\delta(X_1, X_2)}$ 现在需要估计而非设定为 0。同样，$M_1$ 和 $M_2$ 以及 $Y_1$ 和 $Y_2$ 的测量误差也是如此，即 $\theta_{\varepsilon(M_1, M_2)}$ 和 $\theta_{\varepsilon(Y_1, Y_2)}$ 现在需要估计而非设定为 0。这些具有相关性的误差作为参数被估计出来(其方差也要估计，以使协方差可变并且可被估计)。

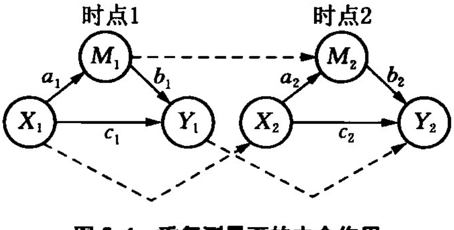



图 5.4 重复测量下的中介作用

拟合如图 5.4 所示的模型的 LISREL 语句如下：

```
Title: mediation among X, Y, Z at times 1&2, and correlated errors be-
tween like terms.
da ni = 6 no = 100 ma = cm
```

```
la
x1 m1 y1 x2 m2 y2
cm sy
1.00
0.30 1.00
0.30 0.30 1.00
0.30 0.30 0.30 1.00
0.30 0.30 0.30 0.30 1.00
0.30 0.30 0.30 0.30 0.30 1.00
se
m1 y1 m2 y2 x1 x2
mo ny = 4 ne = 4 nx = 2 nk = 2 lx = fu, fr td = fu, fr ly = fu, fr te = fu, fr
be = fu, fr ga = fu, fr
pa ga
1 0
1 0
0 1
0 1
pa be
0 0 0 0
1 0 0 0
0 0 0 0
0 0 1 0
pa lx
0 0
0 0
pa td
1 1
```

```
1 1
pa ly
1 0 1 0
0 1 0 1
1 0 1 0
0 1 0 1
pa te
0 0 0 0
0 0 0 0
0 0 0 0
0 0 0 0
va 1.0 lx(1, 1) lx(2, 2) ly(1, 1) ly(2, 2) ly(3, 3) ly(4, 4)
out me = ml rs ef
```

估计出来的参数捕捉了同样的概念建构在不同时间点上进行测量时，测量误差之间的相关性，这对我们来说可能是值得关注的，并且其本身确实有研究价值。例如，这些相关性都接近于 0 就可能意味着样本中的受访者没有刻意使他们的回答在不同的时间点上具有一致性，或者可能表示时点 1 上的测量经历对时点 2 上的测量影响很小，如此等等。典型的情形是，这些估计并不是我们主要的研究兴趣所在，相反，它们被纳入模型只是为了使结构路径系数的估计更纯粹，而后者才真正具有更大的理论意义。

对于结构系数来说，研究者可以用方程 2.4 中的 z 检验同时在时点 1 和时点 2 上对直接路径和间接路径进行比较。此外，还可以使用该方程来量化比较时点 1 和时点 2 上的直接路径。同样，研究者也可以比较两个时间点上的间接路径。然后，研究者会总结得出中介作用的程度是随时间变强了还是变弱了。如果比较不同时间点上的路径系数的 z 检验不显著，就表示中介结构可能没有发生变化，或者是样本规模太小以至于无法探测到某种变化。

# 第 6 章 | 结论

中介分析并非总是必须的。许多具有解释意义的过程都可以从它们所导致的结果中推导出来。如果要进行中介分析，就必须有一个坚实的理论基础，因为相关性数据与得出因果关系结论这一目标之间的关系会掺杂着某种模糊性。如果理论基础足够强，研究者就能够对替代性的模型进行实证检验并提出反对意见。例如，在这些替代性的模型中，变量的角色被交换了或者因果作用的方向被倒转了。

如果要检验中介作用，可以按表 6.1 所给出的步骤进行。这里，我们也简要地对这些步骤进行一些归纳。用结构方程模型拟合一个模型，在这种建模过程中，直接路径和间接路径被同时拟合，目的是在排除了其他影响效应的情况下对每一个影响效应进行估计。当 $X \rightarrow M$ 和 $M \rightarrow Y$ 两个路径系数都显著时，就意味着存在某种程度的中介作用。如果两个路径中有一个不显著(当然也包括两个都不显著这种情况)，分析就此停止，并得出不存在中介作用这一结论。

如果显示出某种程度的中介作用(即间接路径的两段都显著)，并且中介路径的强度大于直接路径，则 z 检验(方程 2.4)将是显著的。如果 z 检验显著并且直接路径 $X \rightarrow Y$ 不显著，则中介作用是完全的。如果 z 检验和直接路径 $X \rightarrow Y$ 都显著，则中介作用是部分的(并且，由 $X$ 所解释的 $Y$ 的变化中，有较大一部分是通过间接路径而非直接路径实现的)。如果 z 检验不显著，但直接路径 $X \rightarrow Y$ 显著，则存在一个直接作用和部分的中介作用(并且，间接路径和直接路径的大小在统计上是相当的)。如果 z 检验和直接路径 $X \rightarrow Y$ 都不显著，则不存在一个直接作用但存在部分的中介作用(并且，间接路径和直接路径的大小在统计上是相当的)。[35]

理想的情形是，每一个概念建构都应该用 3 个或更多的指标变量来测量(Bagozzi & Baumgartner, 1994)。此外，在理想状态下，中介作用应该嵌入一个更大的分析网中，该分析网至少容纳了一个额外的概念建构，并将其作为 $X$ 的一个原因或作为 $X$、$M$ 或 $Y$ 的一个结果。

**表 6.1 用结构方程模型检验中介作用的步骤摘要**

1. 为了检验中介作用，用结构方程模型拟合一个模型，使直接路径和间接路径被同时拟合，目的是在排除了或统计上控制了其他影响效应的同时来估计这两个效应。
   a. 当 $X \to M$ 和 $M \to Y$ 的系数都显著时，意味着存在“一些”中介效应。
   b. 如果两个路径中的任意一个不显著(或者两个都不显著)，就不存在中介作用，研究者应该就此停止。
2. 计算 z 值以明确检验间接(中介)路径与直接路径的相对大小。各种结论如下：
   a. 如果 z 检验显著，但直接路径 $X \to Y$ 不显著，则中介作用是完全的。
   b. 如果 z 检验和直接路径 $X \to Y$ 都显著，则中介作用是部分的(并且，由 $X$ 所解释的 $Y$ 的变化中，有较大一部分是通过间接路径而非直接路径实现的)。
   c. 如果 z 检验不显著，但直接路径 $X \to Y$ 显著(记住间接的中介路径 $X \to M$、$M \to Y$ 是显著的，否则我们就停止分析了)，则存在一个直接作用和部分的中介作用(并且，间接路径和直接路径的大小在统计上是相当的)。
   d. 如果 z 检验和直接路径 $X \to Y$ 都不显著，则不存在一个直接作用但存在部分的中介作用(并且间接路径和直接路径的大小在统计上是相当的)。
3. 研究者可以报告的结果：
   a. 给出路径系数 $a$、$b$、$c$ 的估计值(参数估计和标准误)以及方程 2.4 中的 z 检验。
   b. 报告 SEM 软件包为“间接效应”提供的估计值以及“直接效应”或“总效应”的估计值。
   c. 如果研究者想作出分类性的陈述，他/她可以总结说(根据结果)中介作用是“不存在的”、“部分的”或是“完全的”。
4. 理想的情形是，每一个概念建构都应该用 3 个或更多的指标变量来测量。
5. 核心的三变量中介关系在结构上应该是一个更大的分析网的一部分，该分析网至少容纳了一个额外的概念建构，并将其作为 $X$ 的一个原因或作为 $X$、$M$ 或 $Y$ 的一个结果。
6. 研究者应该承认，有可能存在竞争性的模型，并且要检验几种可能的情形，例如 $M \to X \to Y$，至少要检验 $Y \to M \to X$。理想的情形是加入 $Q$ 来拟合这些竞争性的模型，目的是得到具备诊断性的拟合统计量。然而，即使只有 $X$、$M$ 和 $Y$ 3 个变量，研究者也应该尝试替代性的模型，并且应该能够争辩说，替代性模型中的不同的参数估计比他们中意的模型差。

# 附 录

## 拟合 SEM 中介模型的 SAS、 EQS、 AMOS 命令

**SAS(http://support.sas.com/onlinedoc/913/docMainpage.jsp)**

```
proc calis data = mine cov modification primat toteff;
 * basic mediation;
var x m y; ram
1 3 2 betamy,
1 3 1 gamxy,
1 2 1 gamxm,
2 1 1 phix,
2 2 2 psim,
2 3 3 psiy; run;

proc calis data = mine cov modification primat toteff;
 * mediation with measures: 3 x vars, 3 mediator measures, 2 y measures;
var xl x2 x3 ml m2 m3 y 1 y2; ram
1 1 9 lambxl, 1 2 9 1ambx2, 1 3 9 Iambx3, 1 4 10 lambml, 1 5 10 1ambm2,
1 6 10 lambm 3, 1 7 11 lambyl, 1 8 11 lamdy2, 2 1 1 tdxl, 2 2 2 tdx2, 2 3 3 tdx3,
2 4 4 teml, 2 5 5 tem2, 2 6 6 tem3, 2 7 7 teyl, 2 8 8 tey2,
```

```
1 11 10 betamy,
1 11 9 gamxy,
1 10 9 gamxm,
2 9 9 phix,
2 10 10 psim,
2 11 11 psiy; run;

proc calis data = mine cov modification primat toteff;
 * 4 vars, add q->x;
var x m y q; ram
1 3 2 betamy,
1 3 1 gamxy,
1 2 1 gamxm,
1 1 4 newlink,
2 1 1 phix, 2 4 4 phiq, 2 1 4 phixq,
2 2 2 psim,
2 3 3 psiy; run;

proc calis data = mine cov modification primat toteff;
 * longer chain, x->m1->m2->y;
var x m1 m2 y; ram
1 4 3 betam2y, 1 3 2 bm1m2,
14 1 gamxy,
12 1 gamxm1,
2 1 1 phix,
2 2 2 psim1, 2 3 3 psim2
2 4 4 psiy; run;
```

```
proc calis data = mine cov modification primat toteff;
 * add interaction term;
var x1 x2 x1x2 m y; ram
1 3 2 betamy,
1 4 1 gamx1m, 1 4 2 gamx2m, 1 4 3 gamintm,
1 5 1 gamx1y, 1 5 2 gamx2y, 1 5 3 gaminty,
2 1 1 phix,
2 2 2 psim,
2 3 3 psiy; run;

proc calis data = mine cov modification primat toteff;
 * longitudinal, allow autocorrelations;
var x1 m1 y1 x2 m2 y2; ram
1 3 2 betamy1, 1 6 5 betamy2,
1 2 1 gamxm1, 1 5 4 gamxm2,
1 3 1 gamxy1, 1 6 4 gamxy2,
2 1 4 timex, 2 2 5 timem, 2 3 6 timey,
2 1 1 phix,
2 2 2 psim1,
2 3 3 psiy1, 2 4 4 psix2, 2 5 5 psim2, 2 6 6 psiy2; run;
```

**EQS(http://www.mvsoft.com/products.htm)**

```
/TITLE
My Mediation with Three Constructs, One Measure Each.
/SPECIFICATIONS
CASE = 100; VAR = 3; ME = ML; MA = COV;
/MATRIX
```

```
1.00
0.30 1.00
0.30 0.30 1.00
/LABELS
V1 = X; V2 = M, V3 = Y;
F1 = X; F2 = M; F3 = Y;
/EQUATIONS
F2 = *F1 + D2;
F3 = *F1 + *F2 + D3;
/VARIANCES
D2 = * ; D3 = * ; F1 = * ;
/END

/TITLE
Mediation adding the measurement model, with 3Xs, 3Ms, 2Ys.
/SPECIFICATIONS
CASE = 100; VAR = 8; ME = ML; MA = COV;
/MATRIX
1.00
0.30 1.00
0.30 0.30 1.00
0.30 0.30 0.30 1.00
0.30 0.30 0.30 0.30 1.00
0.30 0.30 0.30 0.30 0.30 1.00
0.30 0.30 0.30 0.30 0.30 0.30 1.00
0.30 0.30 0.30 0.30 0.30 0.30 0.30 1.00
/LABELS
V1 = X1; V2 = X2; V3 = X3; V4 = M1; V5 = M2; V6 = M3; V7 = Y1; V8 = Y2;
```

```
/EQUATIONS
F2 = *F1 + D2;
F3 = *F1 + *F2 + D3;
X1 = *F1 + E1; X2 = *F1 + E2; X3 = *F1 + E3;
M1 = *F2 + E4; M2 = *F2 + E5; M3 = *F2 + E6;
Y1 = *F3 + E7; Y2 = *F3 + E8;
/VARIANCES
D2 = *; D3 = *.F1 = *; E1 = *; E2 = *; E3 = *; E4 = *; E5 = *; E6 =
*; E7 = *; E8 = *;
/END

/TITLE
Four constructs, add q->x.
/SPECIFICATIONS
CASE = 100; VAR = 4; ME = ML; MA = COV;
MATRIX
1.00
0.30 1.00
0.30 0.30 1.00
0.30 0.30 0.30 1.00
/LABELS
V1 = X; V2 = M; V3 = Y; V4 = Q;
F1 = X; F2 = M; F3 = Y; F4 = Q;
/EQUATIONS
F1 = *F4 + D4;
F2 = *F1 + D2;
F3 = *F1 + *F2 + D3;
/VARIANCES
```

```
D2 = * ; D3 = * ; D4 = * ; F1 = * ;
/END

/TITLE
Longer chain, X->M1->M2->Y.
/SPECIFICATIONS
CASE = 100; VAR = 4; ME = ML; MA = COV;
/MATRIX
1.00
0.30 1.00
0.30 0.30 1.00
0.30 0.30 0.30 1.00
/LABELS
V1 = X; V2 = M1; V3 = M2; V4 = Y;
F1 = X; F2 = M1; F3 = M2; F4 = Y;
/EQUATIONS
F2 = *F1 + D2;
F3 = *F2 + D4;
F4 = *F1 + *F3 + D3;
/VARIANCES
D2 = * ; D3 = * ; D4 = * ; F1 = * ;
/END

/TITLE
Add interaction term.
/SPECIFICATIONS
CASE = 100; VAR = 5; ME = ML; MA = COV;
/MATRIX
```

```
1.00
0.30 1.00
0.30 0.30 1.00
0.30 0.30 0.30 1.00
0.30 0.30 0.30 0.30 1.00
/LABELS
V1 = X1; V2 = X2; V3 = X1X2; V4 = M; V5 = Y;
F1 = X1; F2 = X2; F3 = X1X2; F4 = M; F5 = Y;
/EQUATIONS
F4 = *F1 + *F2 + *F3 + D2;
F5 = *F1 + *F2 + *F3 + *F4 + D3;
/VARIANCES
D2 = *; D3 = *; F1 = *; F2 = *;
/END

/TITLE
Longitudinal, as in Figure 5.4.
/SPECIFICATIONS
CASE = 100; VAR = 6; ME = ML; MA = COV;
/MATRIX
1.00
0.30 1.00
0.30 0.30 1.00
0.30 0.30 0.30 1.00
0.30 0.30 0.30 0.30 1.00
0.30 0.30 0.30 0.30 0.30 1.00
/LABELS
V1 = X1; V2 = M1; V3 = Y1; V4 = X2; V5 = M2; V6 = Y2;
```

```
F1 = X1; F2 = M1; F3 = Y1; F4 = X2; F5 = M2; F6 = Y2;
/EQUATIONS
F2 = *F1 + D2;
F3 = *F1 + *F2 + D3;
F5 = *F4 + *F2 + D4;
F6 = *F4 + *F5 + *F3 + D5;
F4 = *F1 + D6;
/VARIANCES
D2 = *; D3 = *; D4 = *; D5 = *; F1 = *;
/END
```

**AMOS(spss.com/amos)**

对大多数人来说，使用 AMOS 软件的好处是它提供了图像化的界面。因此，写出命令在这里反而是多余的。但下面还是给出了拟合基本中介作用的语句：

```
Sub Main
  Dim Sem as New AmosEngine
  'Input File for My Mediation'
  SEM.BeginGroup "C:Myfile.nam"
  Sem.Structure"Y<--M"
  Sem.Structure "Y<--X"
  Sem.Structure "M<--X"
End Sub
```

对于本书所展示的其他模型，在输入数据之后，可使用画图工具绘制出概念建构和变量以及它们之间的连接箭头。

首先,为模拟被纳入中介分析的测量模型,AMOS 的绘图(在点击"run"之前)应该类似于图 3.1(尽管它适用于 3 个 $X$、3 个 $M$ 和 2 个 $Y$ 的情况)

第二,为了模拟有 4 个概念建构的模型(其中 $Q$ 是 $X$ 的一个原因),AMOS 的绘图应该类似于图 4.2(a)。

第三,对一个较长的关系链而言,例如 $X \to M_1 \to M_2 \to Y$,除了这些箭头之外,还应该有 $X \to Y$ 这一直接连接。

第四,为了纳入一个调节变量,应该画出 5 个变量:$X_1$、$X_2$、$X_1X_2$(乘积项)、$M$、$Y$,包括所有从 $X_1$、$X_2$、$X_1X_2$ 到 $M$ 和 $Y$ 的连接以及一个 $M \to Y$ 的连接。

最后,对于图 5.4 所描绘的纵向中介作用,应该有以下连线:$X_1 \to M_1$、$X_1 \to Y_1$、$M_1 \to Y_1$ 以及 $X_2 \to M_2$、$X_2 \to Y_2$、$M_2 \to Y_2$。此外,使用者必须指出自相关连接,即 $X_1 \to X_2$、$M_1 \to M_2$、$Y_1 \to Y_2$。

**MPlus(http://www.statmodel.com/index.shtml)**

```
TITLE:          initial three variable mediation
DATA:           FILE IS my 1.dat;
VARIABLE:       NAMES ARE x m y;
MODEL:          y ON x m;
                m ON x;


TITLE:          mediation with 3 Xs, 3 Ms, 2 Ys
DATA:           FILE IS my2.dat;
VARIABLE:       NAMES ARE x1 - x3 m1 - m3 y1 y2;
MODEL:          f1 BY x1 - x3;
                f2 BY m1 - m3;
                f3 BY y1 - y2;
                y ON x m;
                m ON x;


TITLE:          adding q as fourth construct
```

```
DATA:          FILE IS my3.dat;
VARIABLE:      NAMES ARE x m y q;
MODEL:         y ON x m;
               m ON x;
               x ON q;

TITLE:         longer chain, x to m1 to m2 to y
DATA:          FILE IS my4.dat;
VARIABLE:      NAMES ARE x m1 m2 y;
MODEL:         y ON x m2;
               m2 ON m1;
               m1 ON x;

TITLE:         including a moderator
DATA:          FILE IS my5.dat;
VARIABLE:      NAMES ARE x1 x2 x1x2 m y;
MODEL:         y ON x1 x2 x1x2 m;
               m ON x1 x2 x1x2;

TITLE:         longitudinal
DATA:          FILE IS my6.dat;
VARIABLE:      NAMES ARE x1 m1 y1 x2 m2 y2;
MODEL:         y1 ON x1 m1;
               m1 ON x1;
               y2 ON x2 m2 y1;
               m2 ON x2 m1;
               x2 ON x1;
```

# 注释

[1] Spencer、Zanna 和 Fong(2005)认为，只有当测量中介变量比操控它更容易时，才可能实现。但是请注意，如果对中介变量的测量和对它的操控旨在反映同样的理论概念或建构时，应该不会有什么问题。

[2] 这并不是说研究设计绝不能存在瑕疵。但当研究设计确实有缺陷时，中介作用分析与方差分析一样会存在问题。

[3] Breckler(1990)指出，"横截面数据显然不能证明因果关系推论"，"纵向数据(为确立因果关系证据)提供了更为坚实的基础"。但即便如此，"通常还需要满足其他准则才能使因果关系结论更为稳固"(1990:269)。

[4] Davis(1985:15)曾提到，好的决定因素(例如中介作用中的 $X$)可以是受访者的某些"相对稳定的"特征，包括宗教信仰、职业声望、收入、居住地、政党认同取向、IQ、婚姻状况、子女数等等。好的 $M$ 或 $Y$ 可以是那些"相对不稳定的"或易变的变量，例如幸福感指标、阅读/观看新闻媒介的习惯、品牌偏好和笼统的态度。

[5] 将对操控作用的检验或校对作为 $M$，看似也并不符合通常情况下人们对中介作用的理解和定义。例如，James 和 Brett(1984)这样定义中介作用："如果有概率方程 $M = f(X)$ 和 $Y = f(M)$，那么 $M$ 就在 $X$ 与 $Y$ 之间的概率性关系中充当了一个中介因素，这里，$X$、$M$ 和 $Y$ 必须具有不同的本体论含义(也就是说，它们代表了不同的理论概念或潜在变量)。"(1984:310)

[6] $X \rightarrow Y$ 这一路径在直观上是吸引人的，它回答的是这一问题："$Y$ 的方差是否有某一部分被 $X$ 所解释？如果是，那么这种作用是直接的还是间接的？"然而，1986 年以来，这一规则开始变得有争议。批评的意见认为，中介作用应该是完全的，例如，如果 $X$ 所引起的 $Y$ 的变化全部通过 $M$(或多个 $M$)实现，那么直接路径不显著就是理所当然的。关于这一点的更多讨论，见 James、Mulaik 和 Brett(2007)；Kenny、Kashy 和 Bolger(1998)；Shrout 和 Bolger(2002)。

[7] Sobel(1982)的 z 检验相当重要，因为 Baron 和 Kenny(1986)的步骤(3)被广泛地误解和误用。z 检验要解决的问题是，方程 2.3 中的参数估计值 $c'$(即方程中有 $M$ 时，从 $X$ 到 $Y$ 的直接路径的强度)是否显著地小于方程 2.2(方程中没有 $M$)中的 $c$。目标是要表明 $c'$ 确实"显著地小于"$c$。许多研究者会犯这种错误——他们给出 $c$ 的估计值和相应的 $t$ 值及 $p$ 值(假如，$p$ 值是 0.05)，然后给出 $c'$ 的类似信息(假如，$p$ 值是 0.13)，最后总结说"$c'$ 的显著性小于 $c$"。这种比较与我们所说的毫不相关。第一个参数 $c$ 可能是显著的(即显著地大于 0)，假设 $c = 0.25$，$c'$ 可能是不显著的(即与 0 之间不具有显著的差异)，假设 $c' = 0.10$，但 $c'$ 可能并不显著地小于 $c$。z 检验就是对 0.10($c'$)是否小于 0.25($c$)的一个直接检验。

[8] Baron 和 Kenny(1986)针对"研究总体"在分母中加了一项，使公式变成 $z = (a \times b)/\sqrt{b^2 s_a^2 + a^2 s_b^2 + s_a^2 s_b^2}$。Goodman(1960)针对"样本估计"在分母中减去一项，使公式变成 $z = (a \times b)/\sqrt{b^2 s_a^2 + a^2 s_b^2 - s_a^2 s_b^2}$。由于 $s_a^2$ 和 $s_b^2$ 是方差，其乘积 $s_a^2 s_b^2$ 大于 0，因此，对这两个公式来说，前一个较为保守，后一个更有效。据此，可以轻易地对选择位于两者中间的 Sobel 公式 $z = (a \times b)/\sqrt{b^2 s_a^2 + a^2 s_b^2}$ 进行合理化辩护，而且，它还是一个较好的近似值，尤其是对大样本来说。在任何情况下，参数估计值的方差($s_a^2$ 和 $s_b^2$)通常都较小(例如，0.02)，因此其乘积是可以忽略的(例如，0.004——小数乘小数的值会更小)。中介作用意味着存在一个间接路径。完全的中介作用还意味着直接路径的衰减。如果同时存在一个直接路径和一个间接路径，则意味着部分的中介作用。MacKinnon 等人(2002)曾经指出，这一检验的效力会降低(也就是说，它在某种程度上过于保

守)，他们正在研究如何得出不对称的置信区间，以弥补该估计的不足并对其效力进行提升。

[9] 有人建议，与其将中介分析的结果归纳为这三者之一(即不存在中介作用、部分的中介作用、完全的中介作用)，不如创造出一个连续型的指数来表示间接的中介路径所导致的 $Y$ 的变化比例，即 $ab/(ab+c)$，这种方法所提供的信息会更多(Lehmann, 2001; MacKinnon, Warsi & Dwyer, 1995)。在研究者提出的有关如何正确评估中介作用的文献中，争论之一是，方程 2.1 所界定的情形(即 $X$ 是 $M$ 的一个显著预测变量)意味着，当 $X$ 和 $M$ 被一同用来预测 $Y$ 时(即方程 2.3)，它们之间总是会存在多元共线性。由于这一多元共线性的存在，某些研究者争论说，在概念上以及在统计上，将 $X$ 对 $Y$ 的影响效应和 $M$ 对 $Y$ 的影响效应通过回归权重加以分离是相当困难的，因此，应该比较模型 2.2 和模型 2.3 的拟合性指标 $R^2$，并对它在两个模型之间的变化进行检验(例如，Cote, 2001; Lehmann, 2001)。遗憾的是，也有一些合理的批评意见反对用 $R^2$ 来评估模型之间的比较。更重要的是，模型在总体上解释了多少方差并不能准确地反映哪些变量关系是强的。$R^2$ 这一信息只是对总的模式来说具有诊断性，但却过于宏观以至于不太有用。Baron 和 Kenny (1986)也对多元共线性导致在检验方程 2.3 中的系数时统计效力有所减少表示担心。

[10] Preacher 和 Hayes(2004)尖锐地问，为何对中介作用中的间接作用的显著性检验如此之少？对于不清楚如何进行这种检验的研究者而言，MacKinnon 等人(2002)作出了一个高度综合的介绍(见他们的表 1)。

[11] 见 Mattanah、Hancock 和 Brand(2004)的著作。中介分析是如此常见，以至于它现在已成了荟萃分析的基础(例如，Shadish, 1996)。

[12] 一些研究者正在探索如何使用自助法(一种不对分布作出任何假定的方法，参见 MacKinnon, Lockwood & Williams, 2004)，尤其是对小样本而言(因为基于多元正态性假设的统计量具有渐进性，即适用于非常大的样本，参见 Preacher & Hayes, 2006)。MacKinnon、Warsi 和 Dwyer (1995)发现，对中介效应的估计在有 50 个(或稍多一些)观察个案的样本中表现得相当好，但当使用更大的样本($n>500$)并且自变量被转换为二元取值变量时则不然(当一个连续变量被处理为二元取值变量时，会发生经典的信息损失问题，导致检验变得更保守，这时就需要用大样本来强化数据)。

[13] 另见 www.public.asu.edu/~davidpm/ripl/mediate.htm。

[14] Baron 和 Kenny (1986)写道："一般说来，测量误差的作用是减弱相关性程度，得到的估计值与不存在测量误差的时候相比更接近于 0。例如，只有中介变量上的测量误差会导致自变量对因变量的(直接)作用被高估。"(1986:1177)相反的情形是，如果 $X$ 和 $Y$ 都有测量误差但 $M$ 没有，则中介路径可能会大于直接路径。

[15] Baron 和 Kenny (1986)继续写道："解决信度问题的一般方法是对概念建构进行多种操作或使用多个测量指标……研究者可以使用多元指标这一方法，并用潜变量结构建模方法来估计中介路径。"(1986:1177)

[16] 这一结果可能会使研究者误以为回归方法至少是一种"保守的"方法。然而，随后的结果会表明回归结果是具有误导性的，而 SEM 的结果才接近于研究总体的真实参数 (Iacobucci, Saldanha & Deng, 2007)。这一结果是重要的，因为传统上人们认为，SEM 需要大样本才能表现得较好(例如，$n>200$)。SEM 的缺省计算方法和最常使用的方法是最大似然法，它给出的检验统计量在解释上具有渐进性(即随着 $n$ 的增加，对得到的 $X^2$ 的解释会更加肯定)，因此，问题就是在实践中，$n$ 必须多大才能得到稳定的结果。此外，小样本会在收敛上有问题，其原因与在小样本中用一大堆预测变量来进行回归时会产生杂乱结果的原因相似。该研究表明，$n$ 不需要大到 200，相反，只要 $n=30$ 就足够了。一般来说，该研究所检验的模型并不是非常复杂，但它们肯定涵盖了

中介分析领域。

[17] 第一个下标指的是参数所在矩阵的行，第二个下标指的是列。因此，$\gamma_{MX}$代表了矩阵 Γ 的 $M$ 行 $X$ 列。此外，列变量/概念建构影响了行变量/概念建构，因此，$\gamma_{MX}$代表了 $X \rightarrow M$ 这一影响效应。

[18] LISREL 是这类软件包的鼻祖，其软件语句与标记方法密切相关。此外，其他 SEM 软件包的使用者通常都掌握了 LISREL 的基础工作语言以及 LISREL 与他们中意的软件包之间的转换技术。

[19] 在典型的研究情境下，研究者会尝试创建一个量表以映射某个概念建构，量表的题项与题项之间的相关性会从中等程度到较大程度不等。然而，即使是一个"差"量表，题项之间的相关性也是从较低程度到中等程度不等，把各个题项加在一起也会提高 alpha 值(它是对信度的一个典型测量)。量表也可以是多维的，由于题项之间的相关性相对较高，因此会得到高的 alpha 值，但出于简约性和内部一致性的考虑，理想的情况是使用单维量表，其题项是特别用来表示单一概念建构的。

[20] 与这个更完整的模型相比较，或许之前展示的较为简单的三变量中介模型更清楚。具体来说，情况是这样的：当 3 个概念建构中的每一个都只是由一个变量来测量时，根据方程 3.20，测量与概念建构之间的映射关系就简化为 $\Lambda_X = I$, $\Lambda_Y = 1$, $\Theta_\delta = 0$, $\Theta_\epsilon = 0$(实质上就是假设每一个概念建构在用其指示变量来测量时不存在误差)。大多数中介分析报告都做了这些假设，这有效地将一个本来更大的结构方程模型简化成了一个所谓的路径模型，后者仅仅包含结构模型(而不包含测量模型)。对三变量中介模型而言，方程 3.22 和方程 3.23 就变成了 $Y = I\eta + 0 = \eta$, $X = I\xi + 0 = \xi$. 因此完整模型 3.21 就从 $\eta = \Gamma\xi + B\eta + \Psi$ 简化为 $Y = \Gamma X + BY = \Psi$。

[21] 正是"恰好识别"这一建模问题(自由度为 0、虚假的完全拟合等等)使某些人提出，虽然研究者应该使用结构方程模型，但也应该拟合两个模型来评估可能存在的部分中介作用、一个对应于 $X \rightarrow M \rightarrow Y$ 路径的模型(它只会用掉两个自由度，使剩余的那个可以用来评估拟合度)以及另一个模型(例如，只对应于直接路径 $X \rightarrow Y$ 以进行比较，或对应于所有 3 个路径)。遗憾的是，这种方法延续了拟合多个模型的思路，而在这里，用一个模型同时估计所有参数的方法才具有统计优势。此外，虽然拟合指标会相同，但 3 个核心概念建构之间的关系变动会得出不同的估计(正如我们所展示的那样)，而且某些估计可能在理论上比其他估计更具有合理性。

[22] 如果某个研究者希望 $X$、$M$、$Y$ 所构成的中介关系三角嵌入分析网的方式是 $M$ 或 $Y$ 有一个原因，那么，他就确实应该检验这一模型。理论高于统计，统计只是检验理论的一个工具。估计和检验的过程就像是存在多元共线性的回归分析那样简单，也就是说，对参数估计值的解释必须是更加试探性的或暂时性的，因为彼此相关的概念建构共享了方差。

[23] 含有一个共同因素($M$)既导致了 $X$ 也导致了 $Y$ 这一假设的模型，其估计只基于一个自由度。因此，与拟合饱和模型相比，该模型的拟合将会(典型地)是不完全的，但却是有价值的。

[24] 这些互相对照的情形也阐释了这样一种观点，即 Sobel(1982)的 $z$ 检验，$z = (a \times b)/\sqrt{b^2 s_a^2 + a^2 s_b^2}$，在当路径 $a$ 和路径 $b$ 的强度相当时具有最大的统计效力。例如，作为分子的乘积项在第一种情形下是 $(0.9)(0.15) = 0.135$，在第二种情形下是 $(0.3)(0.89) = 0.267$。只要 $a$ 或 $b$ 中有一个比较小，它们的乘积就会大大减小，即使另一个比较大。

[25] 注意该模型的概念图示并非 $X_1 \rightarrow X_1 * X_2$ 和 $X_2 \rightarrow X_1 * X_2$(加上随后的 $X_1 * X_2 \rightarrow M$ 等等)。虽然这些结构在直观上具有某种吸引力，因为 $X_1$ 和 $X_2$ 在某种意义上对其乘积项/交互项有"贡献"，但是，这已经反映在 $X_1 * X_2$ 这一计算项上，并且已经被包含在协方差矩阵中了。从 SEM 的角度来看，它们并没有"导致"$X_1 * X_2$，而且在 ANOVA 或任何一般的线性模型中，两个主效应项 $X_1$ 和 $X_2$ 及其交互项 $X_1 * X_2$ 是可以独立估计的，并且，在平衡数据中它们都是正交效应。因

此，对 $X_1 \to X_1 * X_2$ 和 $X_2 \to X_1 * X_2$ 进行建模会具有误导性。我们是将 $X_1$、$X_2$、$X_1 * X_2$ 作为单独的外生变量来建模，它们每一个都有指向 $M$ 的路径和指向 $Y$ 的路径。在一个中介性的调节模型中，从主效应开始的路径是作为统计控制的，而从交互项开始的路径才是我们感兴趣的。

[26] Cortina、Chen 和 Dunlap(2001)对这些议题进行了一个概述，Moulder 和 Algina(2002)初步比较了几种可相互替代的分析方法。研究者对预测项之间的多元共线性的担心是存在的，他们也给出了建议，例如在创建交互项之前对主效应变量进行残差对中或均值对中处理(Little, Bovaird & Widaman, 2006)。另外，也存在一些简单的实际方面的问题，例如，当有多个指标时，纳入交互项意味着主效应变量将与每一个指标计算乘积，这会生成许多交互项，而用一个因子得分来计算乘积项则更简单、更符合简约性的要求(Algina & Moulder, 2001; Schumacker, 2002)。最后，Ping(1996a, 1996b, 1996c)建议，如果先完成模型的测量部分，再将测量估计值作为输入常量来拟合结构系数，会得到更好的估计。

[27] Winship 和 Mare(1983)较详细地讨论了以下两种情况的区别：一是对本来就是离散现象的测量(对这种情况我们建议用概率模型，例如，传统的做法是用对数线性模型和 logit 模型来分析这种数据)；二是用离散变量来测量本来是连续型的变量(对这种情况我们一般不明确建议阈值模型，并且，这种数据典型地是用心理计量学和计量经济学方法来分析的)。他们还讨论了这两种情况之间的非线性概率关系(1983: 61—62)。Purists 会争论说，用典型方法收集和分析的评分量表几乎从来都不是真正连续的，例如，不仅仅是从 1 到 7 的评分量表，即使是通过缓慢移动来测量(心理)距离的方法，或者是通过在屏幕上的连续不断的线条上进行标示来“测量”态度的方法，都不是一种真正的连续测量(Bollen, 1989:433)。

[28] 在不同的中介作用强度下(例如，更多或更少的变化是从 $X$ 直接到 $Y$ 发生，或者是间接通过 $M$ 发生)，结果是相似的。在较小的样本规模下，结果也是相似的，虽然有些不稳定。

[29] Bollen(1989:435)报告了一些研究成果，这些研究探讨了积矩相关系数在离散变量情况下的变化。这些研究一致发现，相关(或者基于相关的参数，例如他提到的因子载荷，此外还有推测的路径系数)减弱了(它们会被低估)，并且它们的标准误通常被高估。这些结果强调了那些令人宽慰的经典发现，表明结果至少是更为保守的(并非可能倾向于假设检验中的 I 类错误)。当定类变量的类别数目减少时(比如小于 5)，这种关系强度的弱化效应会加剧。尤其糟糕的情形是在数据分布向相反方向偏斜时，计算两个定类变量之间的相关性。

[30] 在测量理论中，当遇到一个定类变量表示一种潜在的连续型本质时，研究者们表明用这个定类变量来进行替代和替换时，数据中的相关性会减弱，并且将其与多项相关(当一个变量是定序的而另一个变量是连续的)、多分格相关(当两个变量都是定序的)或四分相关(当两个变量都只有二元取值)一起讨论。这些指标有效地估计了基于潜在的连续变量的相关(参见 Bollen, 1989: 442)。这种对数据的预处理可以在 LISREL 中(通过它的附带程序 Prelis)完成。虽然测量理论支持这些校正方法，但期刊审稿人会尤其挑剔，毕竟这些调整使相关性变得更大了。

[31] Fienberg 的原话是：“我们认为赋予数值是有问题的，并且，我们会将自己的结论限定于因果关系的方向指示上。”(1981:121)

[32] 像回归分析一样，我们会说，预测结果是以预测变量为条件的，它们之间的关系在本质上是固定不变的，这意味着这些关系不应该被过度解释，因为它们只是简单地对数据中的样本特征进行了统计控制。然而，$X$ 和 $M$ 之间的关系是可估计的，因此，如果我们愿意假定一种近似性，就可以将 $X$ 和 $M$ 的关系作为 $X \to M$、$M \to Y$ 这一中介连接的一部分。

[33] 报告的统计量是 Wald $X^2$，即$(\beta_{XM}/SE(\beta_{XM}))^2$。如果研究者担心样本较小，可以计算似然比检验统计量作为替代(例如，在 SAS 中说明“Type 3”)。

[34] 广义线性模型就是试图创建一种纲领性的方法来考虑许多不同的模型，类似于"一般线性模型"与ANOVA及回归之间的从属关系。在广义线性模型中，研究者对误差分布的选择（例如，正态、对数正态）以及预测变量和因变量之间的连接函数类型（例如，线性、对数线性）导致了表面上看起来互不相同的技术（例如，回归、logit模型），但这些技术仍然无法实现对两个方程的同步估计，就像这个结构方程：

$$\begin{bmatrix} \hat{M} \\ \hat{Y} \end{bmatrix} = \begin{bmatrix} \gamma_{MX}X + \Psi_M \\ \gamma_{YX}X + \beta_{YM}M + \Psi_Y \end{bmatrix}$$

其中，离散变量 $M$ 会要求连接类型为对数线性、误差分布为对数正态，而连续变量 $Y$ 会要求连接为线性、误差分布为正态。

[35] 研究者可以考虑报告"中介作用的比例"，即，$\hat{a} \times \hat{b} / ((\hat{a} \times \hat{b}) + \hat{c})$。然而，MacKinnon 等人(2002)的研究结果曾表明，这一比率相当不稳定，除非样本量非常大。

# 参考文献


Algina, J. & Moulder, B. C. (2001). "A note on estimating the Joreskog-Yang model for latent variable interaction using LISREL 8. 3." *Structural Equation Modeling 8*(1):40—52.

Asher, H. B. (1983). *Causal models*(2nd ed.). Beverly Hills, CA: Sage.

Bagozzi, R. P. & Baumgartner, H. (1994). "The evaluation of structural equation models and hypothesis testing." In R. P. Bagozzi(ed.), *Principles of marketing research*. Cambridge, MA: Blackwell.

Baron, R. M., Kenny, D. A. (1986). "The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations." *Journal of Personality and Social Psychology 51*(6):1173—1182.

Baumrind, D. (1983). "Specious causal attributions in the social sciences." *Journal of Personality and Social Psychology 45*(6):1289—1298.

Bentler, P. (2001). "Mediation." *Journal of Consumer Psychology 10*(1/2):84.

Bollen, K. A. (1987). "Total, direct, and indirect effects in structural equation models." *Sociological Methodology 17*:37—69.

Bollen, K. A. (1989). *Structural equations with latent variables*. New York: Wiley.

BrecMer, S. J. (1990). "Applications of covariance structure modeling in psychology: Cause for concern?" *Psychological Bulletin 107*(2):260—273.

Brown, R. L. (1997). "Assessing specific mediational effects in complex theoretical models." *Structural Equation Modeling 4*(2):142—156.

Churchill, G. A., Jr. & Iacobucci, D. (2005). *Marketing research: Methodological foundations* (9th ed.). Mason, OH: Thomson Learning.

Cole, D. A. & Maxwell, S. E. (2003). "Testing mediational models with longitudinal data: Questions and tips in the use of structural equation modeling." *Journal of Abnormal Psychology 112*(4):558—577.

Collins, L. M., Graham, J. W. & Flaherty, B. P. (1998). "An alternative framework for defining mediation." *Multivariate Behavioral Research 33*(2):295—312.

Cortina, J. M., Chen, G. & Dunlap, W. P. (2001). "Testing interaction effects in LISREL: Examination and illustration of available procedures." *Organizational Research Methods 4*(4):324—360.

Cote, J. (2001). "Mediation." *Journal of Consumer Psychology 10*(1/2):93—94.

Cronbach, L. & Meehl, P. (1955). "Construct validity in psychological tests." *Psychological Bulletin 52*(4):281—302.

Davis, J. A. (1985). *The logic of causal order*. Beverly Hills, CA: Sage.

DiStefano, C. (2002). "The impact of categorization with confirmatory factor analysis." *Structural Equation Modeling 9*(3):327—346.

Ethington, C. A. (1987). "The robustness of LISREL estimates in structural equation models with categorical variables." *Journal of Experimental Education 55*(2):80—88.

Fienberg, S. E. (1981). *The analysis of cross-classified categorical data* (2nd ed.). Cambridge: MIT Press.

Glymour, C. (2001). *The mind's arrows: Bayes nets and graphical causal models in psychology.* Cambridge: MIT Press.

Goodman, L. (1960). "On the exact variance of products." *Journal of the American Statistical Association 55*:708—713.

Haughton, D., Kamis, A. & Scholten, P. (2006). "A review of three directed acyclic graphs software packages: MIM, Tetrad, and WinMine." *The American Statistician 60*(3):272—286.

Hayduk, L., Cummings, G., Stratkotter, R., Nimmo, M., Grygoryev, K., Dosman, D., et al. (2003). "Pearl's D-Separation: One more step into causal thinking." *Structural Equation Modeling 10*(2):289—311.

Holbert, R. L. & Stephenson, M. T. (2003). "The importance of indirect effects in media effects research: Testing for mediation in structural equation modeling." *Journal of Broadcasting & Electronic Media. December*:556—572.

Holland, P. W. (1986). "Statistics and causal inference." *Journal of the American Statistical Association 81*(396):945—960.

Holland, P. W. (1988). "Causal inference, path analysis, and recursive structural equations models." In C. C. Clogg(ed.), *Sociological methodology* (Vol. 18, pp. 449—484). Washington, DC: American Sociological Association.

Humphreys, P. & Freedman, D. (1996). "The grand leap." *British Journal of the Philosophy of Science* 47:113—123.

Iacobucci, D. (2001). "Special issue on the methodological questions of the experimental behavioral researcher." *Journal of Consumer Psychology 10*(1/2):83—100.

Iacobucci, D., Saldanha, N. & Deng, J. X. (2007). "A meditation on mediation: Evidence that structural equations models perform better than regressions." *Journal of Consumer Psychology 17*(2):140—154.

James, L. R. & Brett, J. M. (1984). "Mediators, moderators, and tests for mediation." *Journal of Applied Psychology 69*(2):307—321.

James, L. R., Mulaik, S. A. & Brett, J. M. (1982). *Causal analysis: Assumptions, models, and data.* Beverly Hills, CA: Sage.

James, L. R., Mulaik, S. A. & Brett, J. M. (2007). "A tale of two methods." Manuscript under review.

Jöreskog, K. & Sörbom, D. (1997). *LISREL 8: User's Reference Guide.* Chicago: SSI Scientific Software.

Judd, C. M., Kenny, D. A. & McClelland, G. H. (2001). "Estimating and testing mediation and moderation in within-subject designs." *Psychological Methods 6*(2):115—134.

Karson, E. J. & Fisher, R. J. (2005). "Reexamining and extending the dual mediation hypothesis in an on-line advertising context." *Psychology & Marketing 22*(4):333—351.

Kenny, D. A. Kashy, D. A. & Bolger, N. (1998). "Data analysis in social psychology." In D. Gilbert, S. T. Fiske & G. Lindzey(eds.), *Handbook of social psychology* (Vol. 1, pp. 233—

265). New York: McGraw-Hill.

Kline, R. B. (1998). *Principles and practice of structural equation modeling*. New York: Guilford Press.

Knoke, D. & Burke, P. J. (1980). *Log-linear models*. Beverly Hills, CA: Sage.

Kupek, E. (2005). "Log-linear transformation of binary variables: A suitable input for SEM." *Structural Equation Modeling 12*(1):28—40.

Lee, S. & Hershberger, S. (1990). "A simple rule for generating equivalent models in covariance structure modeling." *Multivariate Behavioral Research 25*(3):313—334.

Lehmann, D. (2001). "Mediation."*Journal of Consumer Psychology 10*(1/2):90—92.

Little, T. D., Bovaird, J. A. & Widaman, K. A. (2006). "On the merits of orthogonalizing powered and product terms: Implications for modeling interactions among latent variables." *Structural Equation Modeling 13*(4):497—519.

Long, J. S. (1997). *Regression models for categorical and limited dependent variables*. Thousand Oaks, CA: Sage.

Long, J. S. (2006). *Covariance structure models: An introduction to LISREL*. Thousand Oaks, CA: Sage.

Lynn, H. S. (2003). "Suppression and confounding in action." *The American Statistician 57*(1): 58—61.

MacCallum, R. C., Wegener, D. T., Uchino, B. N. & Fabrigar, L. R. (1993). "The problem of equivalent models in applications of covariance structure analysis." *Psychological Bulletin 114*(1):185—199.

Mackie, J. L. (1974, 1980). *The cement of the universe: A study of causation*. Oxford, UK: Clarendon Press.

MacKinnon, D. P., Krull, J. L. & Lockwood, C. M. (2000). "Equivalence of the mediation, confounding and suppression effect." *Prevention Science 1*(4):173—181.

MacKinnon, D. P., Lockwood, C. M., Hoffman, J. M., West, S. G. & Sheets, V. (2002). "A comparison of methods to test mediation and other intervening variable effects." *Psychological Methods 7*(1):83—104.

MacKinnon, D. P., Lockwood, C. M. & Williams, J. (2004). "Confidence limits for the indirect effect: Distribution of the product and resampling methods." *Multivariate Behavioral Research 39*(1):99—128.

MacKinnon, D. P., Warsi, G. & Dwyer, J. H. (1995). "A simulation study of mediated effect measures." *Multivariate Behavioral Research 30*(1):41—62.

Mattanah, J. F., Hancock, G. R. & Brand, B. L. (2004). "Parental attachment, separation-individuation, and college student adjustment: A structural equation analysis of meditational effects." *Journal of Counseling Psychology 51*(2):213—225.

McDonald, R. (2001). "Mediation." *Journal of Consumer Psychology 10*(1/2):92—93.

McDonald, R. (2002). "What can we learn from the path equations? Identifiability, constraints, equivalence." *Psychometrika 67*(2):225—249.

McKim, V. R. & Turner, S. P. (1997). *Causality in crisis? Statistical methods and the search for*

*causal knowledge in the social sciences*. South Bend, IN: Notre Dame Press.

Menard, S. (1995). *Applied logistic regression analysis*. Thousand Oaks, CA: Sage.

Mill, J. S. (2002). *A system of logic: Ratiocinative and inductive*. Honolulu, HI: University Press of the Pacific.

Moulder, B. C. & Algina, J. (2002). "Comparison of methods for estimating and testing latent variable interactions." *Structural Equation Modeling* 9(1):1—19.

Muller, D., Judd, C. M. & Yzerbyt, V. Y. (2005). "When moderation is mediated and mediation is moderated." *Journal of Personality and Social Psychology* 89(6):852—863.

Muthén, B. (1984). "A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators." *Psychometrika* 49(1):115—132.

Netemeyer, R. (2001). "Mediation." *Journal of Consumer Psychology* 10(112):83—84.

Pearl, J. (2000). *Causality: Models, reasoning, and inference*. Cambridge, UK: Cambridge University Press.

Pearl, J. (2001). "Direct and indirect effects." *Proceedings of the 17th conference on Uncertainty in Artificial Intelligence*. San Francisco: Morgan Kaufmann.

Ping, R. A., Jr. (1996a). "Estimating latent variable interactions and quadratics: The state of this art." *Journal of Management* 22(1):163—183.

Ping, R. A. Jr. (1996b). "Latent variance interaction and quadratic effect estimation: A two-step technique using structural equation analysis." *Psychological Bulletin* 119(1):166—175.

Ping, R. A., Jr. (1996c). "Latent variance regression: A technique for estimating interaction and quadratic coefficients." *Multivariate Behavioral Research* 31(1):95—120.

Preacher, K. J. & Hayes, A. F. (2004). "SPSS and SAS procedures for estimating indirect effects in simple mediation models." *Behavior Research Methods, Instruments, & Computers* 36(4):717—731.

Preacher, K. J. & Hayes, A. F. (2006). "Asymptotic and resampling strategies for assessing and comparing indirect effects in simple and multiple mediator models." Unpublished manuscript.

Rubin, D. B. (1974). "Estimating causal effects of treatments in randomized and nonrandomized studies." *Journal of Educational Psychology* 66(5):688—701.

Rubin, D. B. (2005). "Causal inference using potential outcomes: Design, modeling, decisions." *Journal of the American Statistical Association* 100(469):322—331.

Salmon, W. C. (1997). *Causality and explanation*. Oxford, UK: Oxford University Press.

Schumacker, R. E. (2002). "Latent variable interaction modeling." *Structural Equation Modeling* 9(1):40—54.

Schumacker, R. E. & Beyerlein, S. T. (2000). "Confirmatory factor analysis with different correlation types and estimation methods." *Structural Equation Modeling* 7(4):629—636.

Shadish, W. R. (1996). "Meta-analysis and the exploration of causal mediating processes: A primer of examples, methods, and issues." *Psychological Methods* 1(1):47—65.

Shafer, G. (1996). *The art of causal conjecture*. Boston: MIT Press.

Shipley, B. (2000). "A new inferential test for path models based on directed acyclic graphs." *Structural Equation Modeling* 7(2):206—218.

Shrout, P. E. & Bolger, N. (2002). "Mediation in experimental and nonexperimental studies: New procedures and recommendations." *Psychological Methods* 7(4):422—445.

Simon, H. (1957). "Spurious correlations: A causal interpretation." In *Models of man: Social and rational*. New York: Wiley.

Sobel, M. E. (1982). "Asymptotic confidence intervals for indirect effects in structural equation models." In S. Leinhardt(ed.), *Sociological methodology*. San Francisco: Jossey-Bass.

Sosa, E. & Tooley, M. (1993). *Causation*. Oxford, UK: Oxford University Press.

Spencer, S. J., Zanna, M. P. & Fong, G. T. (2005). "Establishing a causal chain: Why experiments are often more effective than mediational analyses in examining psychological processes." *Journal of Personality and Social Psychology* 89(6):845—851.

Spirtes, P., Glymour, C. & Scheines, R. (2001). *Causation, prediction, and search: Adaptive computation and machine learning* (2nd ed.). New York: Bradford Books.

Suppes, P. (1970). *A probabilistic theory of causality*. Amsterdam: North-Holland.

Stelzl, I. (1986). "Changing a causal hypothesis without changing the fit: Some rules for generating equivalent path models." *Multivariate Behavioral Research 21*:309—331.

Tzelgov. J. & Henik, A. (1991). "Suppression situations in psychological research: Definitions, implications, and applications." *Psychological Bulletin 109*(3):524—536.

Winship, C. & Mare, R. D. (1983). "Structural equations and path analysis for discrete data." *American Journal of Sociology 89*(1):54—110.

# 多层次模型

## 作 者 简 介

### 道格拉斯·A. 卢克(Douglas A. Luke)

美国圣路易斯大学公共健康学院副教授、生物统计学系主任。其研究领域主要包括烟草控制评估与政策以及与健康行为相关的媒体政策。

## 译 者 简 介

### 郑冰岛

曾先后获得中国人民大学社会学系法学学士学位、硕士学位。2010 年获香港科技大学社会科学部研究型硕士学位。现为香港科技大学社会科学部博士研究生,研究领域为社会分层和流动、政治社会学。

# 序

在研究人类行为的模型中，"情境"具有非同寻常的重要性。个体行为受到多个层次自变量的影响，可以是微观层面的，也可以是宏观层面的。例如，某投票者对候选人捐赠的竞选资金是个体特征（投票者的收入和教育）与群体特征（投票者的社区或其所在的专业机构）的共同函数。在该情况下，研究者可以估计一个 OLS 模型，$V = a + bI + cS + dN + eP + u$。其中，V、I 和 S 分别表示个体层面的捐赠值、收入以及教育；N 和 P 则分别表示群体层面的社区和专业机构特征；b 和 c 相应为个体效应的参数；d 和 e 为群体效应参数；u 为误差项。这是一种有效的估计方法，在 Iversen 的《情境分析》一书中，对此方法有详尽的讨论。

但出现多层次效应时，OLS 的经典假设却很难满足。特别是当个体存在于相同群体中时，"不相关的误差项"的假设几乎不可能成立。本书采用最大似然方法估计多层次模型来解决这一问题。多层次模型还有一些其他的名称，如分层线性模型、随机系数模型、混合效应模型，并且它可以使用单个方程，亦可运用联列方程的形式。而 Luke 博士关注单个方程回归形式的二层到三层模型。通过讨论国会成员对烟草工业的投票行为这一例子，Luke 博士解释了个体层面及集体层面特征共同作用于同一函数的例子，为二层模型提供了最大似然估计（MLE）方法，并且对模型的拟合优度指标进行了详细的阐释，包括离差、赤池信息准则（AIC）和贝叶斯信息准则（BIC）等。

本书还对该技术的使用进行了扩展讨论。例如讲述离散变量和非正态分布变量的一般分层线性模型（GHLM）；又如使用健康调查数据讨论不同时点嵌套于同一个体的纵向数据分析，其中受访者在多个时间点上接受访谈，且不同个体间数据的缺失情况有很大差异。这类情况在标准的重复测量多变量方差分析（MANOVA）中很难处理，但使用多层次模型即可解决这一问题。纵向数据中常见的自相关问题也被讨论过。由于多层次模型的设计和分析方法都处在日新月异的进展之中，因而本书还介绍了一系列新的应用软件，主要是现在流行的两个专业软件包——HLM 和 R/S-Plus。同时，本书还介绍了一些更普遍的程序，如 SAS 和 SPSS。若研究者希望阅读一些以平实的语言对多层次模型进行分析的著作，本书即其中具有代表性的一本。

迈克尔·S. 刘易斯-贝克

# 第1章 | 为什么使用多层次模型

在考虑社会与健康科学所关注的几乎所有现象时，我们都应该注重事件情境的重要性。例如，社会环境的紧张可能影响个体的情绪；药物对心理状态的改善受使用者所处社会结构的影响；早期童年发展与一系列环境条件相关，如饮食习惯、环境刺激、环境污染、与母亲的关系等等；青少年危险行为和他们所处的成人世界的结构化行为有关系；孩子的教育成果会被课堂、学校以及教学体制所制约。

以上都是个体行为受到社会环境影响的例子。实际上，情境的作用远不止这些。例如，婚姻关系中离婚的选择与宗教和文化背景显著相关；组织氛围影响群体选择；医疗机构利润受卫生维护组织（HMO）返款政策的强烈影响。

上述所有例子的共同点在于，某一个高层次特征的存在或发生过程影响着低层次特征的存在或发生。由于概念是建构于不同层次的，因此概念间的关系将发生在多层次之间，从而需要特殊的分析工具才能加以估计。这就是本书的主题。

尽管情境如此重要，但那些并不能解决多层次数据与理论问题的分析工具却常被运用于健康与社会科学研究中。在研究发展的早期阶段，这是由于多层次分析工具的缺失造成的。然而在多层次模型已经相当完善的今天，研究者却继续使用简单化的单一层次模型。

社会科学中某些传统的认识论或许可以解释这种现象。首先是实证主义研究传统的广泛影响，即便是在多年以前，科学哲学已经认识到实证主义并不足以建立生物、健康以及社会科学的研究框架，我们却倾向于追随实证主义传统来进行研究设计和分析。例如在研究中，我们会控制实验条件、使用控制组

和对照组以及在模型中控制协变量的影响以提供更准确的统计推论，但在测量和估计超个体层面的环境性因素时，我们却面临非常严重的局限。

实证主义在处理科学问题时最为有效，因其面对的大多是封闭体系中的问题。这类封闭体系中的行为常常可以用少数几个变量来预测，比如行星的位移可用其质量和速度来估计。健康与社会科学处理的是更为复杂的开放体系中的问题，其中外部环境因素常常无法被控制、限定或去除，从而使能够测量和分析环境因素的多层次模型显得尤为重要。

让我们来看一个医学模型的例子。它以简化的视角来看待健康，将疾病视为可被药物刺激而修正的一种身体缺陷，但现代流行病学则试图确定疾病的风险因素。以心血管疾病为例，现代流行病学通过一系列有力的研究设计（如控制组研究）和分析工具（如对相对患病风险的 logistic 回归）来确定重要的患病风险因素，包括基因体质、身体状况（高血压）、行为习惯（吸烟、运动）、文化背景（种族）和环境（医疗条件）。然而，虽然这些因素明显作用于不同层面，但它们却总是于个体层面被测量（比如通过调查），并且各不同层面的因素发生作用的机制也很少被关注。比如，缺乏运动是个体选择造成的，还是由于社区生态导致个体在社区内缺乏运动条件造成的？

尽管如此，在行为、健康和社会科学中，多层次模型却正在获得越来越多的关注和使用。在美国国立卫生研究院 2000 年的报告《走向多层次分析：健康研究中社会和文化维度的进步与前景》（行为及社会科学研究办公室，2000）中，这份提出了新议题的报告的研究目标有两个：一是扩展与健康相关的社会科学研究；二是将社会科学研究整合进跨学科的、多层次的健康研究中。为了实现这两个目标，报告提出了以下建议：

> 支持发展高质量的社会科学方法。这方面的挑战包括：群体、网络、邻里、社区层次的测量；纵向数据研究方法的发展；整合不同量化与质化研究的多层次研究设计以及数据搜集和分析方法的改进。

表 1.1 以烟草控制为例，这份报告展示了各个分析层次之间的相互影响，并

以此构建概念框架。烟草控制研究存在于从基因到社会文化、政治的各个层面。尽管可以将研究局限于任何一个层面，但大多数重要的研究会关注各个层面之间的联系。例如，在了解各种基因的尼古丁依赖的基础上，我们可以对各基因类型进行恰当的预防性干预。

**表 1.1　健康研究的不同层次：以烟草控制为例**

| 分析层次 | 示例：烟草控制研究 |
| --- | --- |
| 文化/政治 | 测量烟草税对人口吸烟比例的影响 |
| 社会/环境 | 测量家庭与同辈影响在青少年吸烟行为中的相对重要性 |
| 行为/心理 | 设计有效的吸烟预防与禁止项目 |
| 器　官 | 防止吸烟者体内肿瘤的形成 |
| 细　胞 | 尼古丁摄入的新陈代谢研究 |
| 分子/基因 | 尼古丁依赖的基因研究 |

2003 年美国国家科学院医学研究所关于公众健康的报告更清楚地反映了多层次因素的相互依赖性和层级化特征。图 1.1 展现了健康决定因素的社会生态模型。该报告强调，公共健康专家以及研究者必须理解和应用社会生态学路径，以期成功地改善国家整体的健康状况。

**图 1.1　影响健康的决定因素的多层次索引图**

## 多层次模型的理论依据

简单地说，由于大多数研究对象在本质上是多层次的，因而我们应使用多层次的理论和分析工具，否则我们将面临严重的问题。

例如，在集体层面搜集和分析健康行为数据是非常普遍的。流行病学显示，在那些居民日常食谱中脂肪含量较高的国家，其乳腺癌致死的比例也较高(Carroll, 1975)。如此看来，摄入更多脂肪的女性更易患乳腺癌似乎是一个合理的推论。然而这一推论却是一个生态学谬误，其中，群体间的关系被假设同样存在于个体之间(Freedman, 1999)。事实上，近期的健康研究表明，在个体层面，脂肪摄入量与乳腺癌之间的关系并不明显(Holmes et al. , 1999)。

这类谬误也可能以另一种方式出现。在行为科学中，数据常常从个体处搜集，然后集合成为群体数据以获得个体所从属的群体信息，这可能导致原子谬误，即群体信息被不正确地从个体信息中推论出来(Hox, 2002)。尽管从个体层面获取整个生态的特征信息并非不可能，如 Moos 关于社会气候的推论就是一个成功的例子(Moos, 1996)，然而 Shinn 和 Rapkin 却令人信服地指出，这类推论并不可靠，更可取的方法是使用群体层次的测量和分析工具来获取群体信息。

我们有必要考虑集群和成员在属性上的社会学区别(Lazarsfeld & Menzel, 1969)。成员从属于集群，但集群与其成员的多种属性(变量)却可能同时被测量和分析。Lazsarfeld 和 Menzel 定义了集群的分析属性、结构属性以及整体属性。分析属性通过集合集群内的个体信息来获得(例如，一个城市中西班牙裔人口的比例)，结构属性建立在集群成员之间的关系上，而整体属性则是集群本身的特征，并不受个体成员的影响(O'Brien, 2000)。例如，学校的禁烟政策就是学校这一集群的整体属性。

从这一框架可以看出，以上的谬误存在于推论而非测量中。我们完全可以从低维度的成员信息中获得高层次的集群特征。当某一特定层面的关系被不恰当地假定于其他层面时，上述谬误就出现了。

## 多层次模型的统计依据

当面对这些复杂的概念问题时，即使所处理的数据与假设实为多个层次的，社会科学研究者仍倾向于使用传统的个体层次模型来解决问题。若分解群体层次的信息至个体层面，从而将多元回归中的变量限定在个体分析单位中，则至少会导致两方面的问题：首先，所有未被拟合入模型的背景信息最终都被包含在模型的个体层次误差项中（Duncan, Jones & Moon, 1998），而由于相同背景下的个体之误差项必然相关，则违反了多元回归的基本假设；其次，忽略背景因素则意味着各回归系数同等作用于一切情境，这实则反映了"在不同背景条件下，事物的发生机制本质相同"的错误观点（Duncan et al. , 1998:98）。

**表 1.2 健康和社会科学研究中的多层次模型与结构**

| 多层次模型 | 多层次结构 | 示　例 |
|---|---|---|
| 物　理 | 个体存在于物理环境之中，这包括生物环境、生态环境以及物理建构环境 | Diez-Roux 等人(2001)<br>Perkins 等人(1993) |
| 社　会 | 个体存在于社会结构之中，包括家庭、同辈以及其他社会网络 | Buka 等人(2003)<br>Rice 等人(1998) |
| 组　织 | 个体和小群体存在于特定的组织之中，重要的组织特征包括规模、管理结构、群体交流、组织目标等 | Maes 与 Lievens(2003)<br>Villemez 与 Bridges(1998) |
| 政治/文化 | 个体或群体存在于特定的社会政治环境、文化环境和历史环境中 | Lochner 等人(2001)<br>Luke 与 Krauss(2004) |
| 时　间 | 对某一个体在不同时点的多次观察 | Boyle 与 Willms(2001)<br>Curran、Stice 与 Chassin(1997) |
| 分　析 | 个体研究中的多元效应测量（如元研究） | Goldstein 等人(2000)<br>Raudenbush 和 Bryk(1985) |

通过引入个体分组情况的影响，方差分析（ANOVA）与协方差分析（ANCOVA）可以部分地解决这些统计问题，但缺陷仍然存在：首先，当个体的组别较多时，这类模型必然使用较多的参数而大大降低解释力和简约性；其次，这些组别参数往往被作为固定效应引入，从而忽略了群组层面的随机变量；最后，在处理缺失数据或不均衡设计时，方差分析不够灵活。

## 本书内容简介

上述讨论的主要目的在于对社会与健康科学研究中的多层次模型统计方法做相对非技术性的介绍。之后的内容分为两大部分。第 2 章介绍二层模型及其拟合方法，包括数据准备、模型估计、模型解释、假设检验、模型假设条件检验以及中心化。第 3 章是对多层次模型的扩展应用，包括对非连续型因变量和非正态分布型因变量的处理以及使用多层次方法分析纵向数据和构建三层模型。这些内容都讨论了其与多元回归的相似性，并且对例子中的数据和分析进行了扩展运用(所有的数据、程序以及分析结果都可在附录中找到)。而本书所有的数据分析都使用了 HLM5. 04 版(Raudenbush, Bryk, Cheong, Congdon, 2000)以及 R 程序(Pinheiro, Bates, DebRoy, Sarkar, 2003)里的混合效应语句 nlme3. 1 版，所有的图都由 R1. 7. 1 版产生。

为了方便下文的叙述，我们先对多层次模型提供一个定义，它适用于多层次数据以阐述不同层级间关系的统计技术。在过去的几十年间，多层次模型的统计基础在各学科内发展起来，并被给予不同的称谓，包括分层线性模型(Raudenbush & Bryk, 2002)、随机系数模型(Longford, 1993)、混合效应模型(Pinheiro & Bates, 2000)、协方差结构模型(Muthén, 1994)以及增长曲线模型(McArdle & Epstein, 1987)等。所有这些多层次模型的具体形式都可归纳为以下两大统计类别中的一种：多元回归统计和结构方程模型。本书关注前一种统计路径，读者可以参见 Heck 和 Thomas 的著作(Heck & Thomas, 2000)第 5 章至第 7 章的内容，借此对第二种统计路径有所了解。

# 第 2 章 | 基本多层次模型

## 基本二层模型

多层次模型的目的在于基于一系列非同一层次的自变量，对因变量的值进行估计。例如，对儿童标准化阅读能力的估计不仅要考虑儿童的个体特征（如学习时间），亦应考虑儿童所处班级的集体特征（如班级规模）。在此，我们将儿童特征置于测量和模型化的第一层，而将班级特征置于第二层。

使用分处第一层和第二层的两个自变量，上述二层结构可以写成下列的多层次模型：

$$
\begin{aligned}
\text{第一层：} Y_{ij} &= \beta_{0j} + \beta_{1j}X_{ij} + r_{ij} \\
\text{第二层：} \beta_{0j} &= \gamma_{00} + \gamma_{01}W_j + u_{0j} \\
\beta_{1j} &= \gamma_{10} + \gamma_{11}W_j + u_{1j}
\end{aligned}
\qquad [2.1]
$$

这一组方程不仅列出了所有因变量和自变量，还更清晰地描述了模型的多层次特征。模型第一层看似一个典型的 OLS 多元回归，然而下标 $j$ 却表明其估计随着第二层特征（班级）值的变化而不同，即研究对象中的每个班级都有其各自的平均阅读能力（$\beta_{0j}$），而且学习时间对阅读能力的影响（$\beta_{1j}$）也有所不同。因而我们允许模型的截距和斜率在不同的班级之间变化，这也是多层次模型的核心观点——将截距和斜率作为第二层自变量的结果。

模型的第二层指出了第一层模型的参数如何受到第二层变量的影响：$\beta_{0j}$ 是第一层模型的截距；$\gamma_{00}$ 是在控制二层自变量 $W_j$ 时，一层因变量的均值；$\gamma_{01}$ 是二层自变量 $W_j$ 的斜率；$u_{0j}$ 是误差项，即未被模型化的变项。第二个方程的解

释与此类似，但此方程拟合的是二层变量对 $X_{ij}$ 的斜率。$\beta_{1j}$ 是一层模型的斜率，$\gamma_{10}$ 是控制 $W_j$ 时的均值，$\gamma_{11}$ 是 $W_j$ 的斜率，而 $u_{1j}$ 是其误差项。

将第二层方程代入第一层可得：

$$Y_{ij} = \underbrace{[\gamma_{00} + \gamma_{10}X_{ij} + \gamma_{01}W_j + \gamma_{11}W_jX_{ij}]}_{\text{固定}} + \underbrace{[u_{0j} + u_{1j}X_{ij} + r_{ij}]}_{\text{随机}} \quad [2.2]$$

此为多层次模型的单个方程形式，又名“混合效应模型”。它相对简洁，但却难以迅速辨别其隐含模型的多层次结构。其优点在于，首先，如上所叙，单个方程清楚地表明了模型的固定效应部分($\gamma$)以及随机效应部分($u$ 和 $r$)，这也是为何多层次模型亦被称为“混合模型”或“混合效应模型”的原因；另外，单个方程形式与多层次模型软件的输出结果相对应。它表明第一层次的参数（$\beta_{0j}$、$\beta_{1j}$）并不被直接估计，而是通过第二层参数($\gamma$)来间接获得。

大多数研究者，尤其是方差分析方法的使用者，都对固定效应较为熟悉，对随机效应则不然。在方差分析中，随机效应被定义为一些独立因素，其所处的层级是随机选择的。而在多层次模型中，随机效应指的是附加误差项或是额外的方差。在上文的例子中，$\gamma_{ij}$ 是个体层次的误差项，而多层次模型还有两项附加的误差项：$u_{0j}$ 是各个班级阅读水平的随机差异，$u_{1j}$ 则是各个班级学习时间与阅读水平关系的随机差异。由此看来，多层次模型的随机效应总是依赖于各个层级单位。

方程 2.1 是一种相当典型的多层次模型形式。多层次模型还有许多其他形式，我们一般依据实际情况选择某一形式进行估计。为了处理可能模型的多种让人困惑的组合，我们根据以下 4 个方面来决定最终模型：

第一，现有几层数据？其中有几层将被模型化？尽管三层以上的模型并非不可能，但现有社会科学研究文献中常常使用二层到三层模型。

第二，各层次分别有多少自变量需加以考虑？

第三，是否将一层的斜率或截距，或两者共同作为二层特征的结果？图 2.1 展示了其中的不同。图 2.1 左侧反映了斜率不变而截距在不同二层单位间变动的情况，而右侧则是斜率和截距都发生变动时的情况。研究者应综合考虑理论

背景及数据证据来作出决定。

第四，模型的哪部分需引入随机效应？

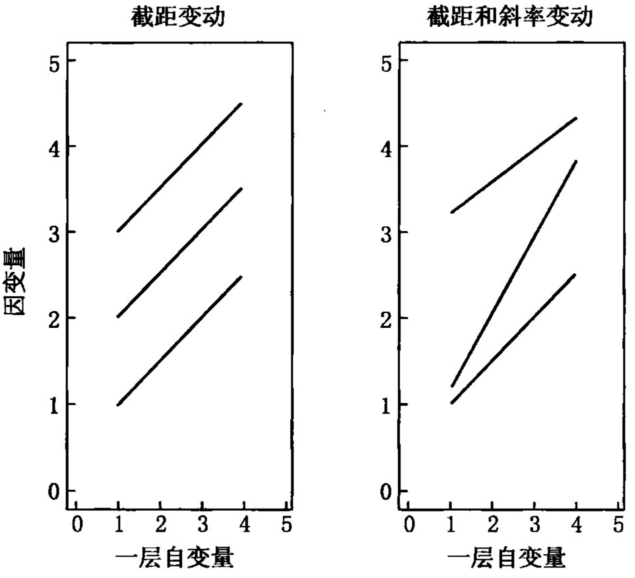



图 2.1 截距变动及截距与斜率共变的例子

尽管多层次模型的形式数不胜数，但我们仍大致将其分为 3 类，见表 2.1。第一类是多层次模型中的最简形式，不含有任何一层或二层的自变量，被称为“非限制模型”或“原模型”。它常被作为起点，用以构建更复杂的模型，特别是用于计算组间相关系数(如下所示)。

第二类模型假设截距在二层单位间变化而斜率不变。在我们的例子中，若我们相信不同班级的平均阅读水平不同，但个体学习时间对阅读能力的影响在各班之间没有变化时，则可使用这类模型。

最后一类模型假设截距和斜率共同变化。若我们相信班级特征与个体学习时间对阅读能力的影响的层间交互作用存在，则可以用这类模型。例如，有些教师倾向于给学习勤奋的学生较好的阅读成绩，但有些教师却并不以学生的学习时间为衡量准则。方程 2.2 中的 $\gamma_{11}W_jX_{ij}$ 即代表这一层间交互作用。

**表 2.1 3类多层次模型**

| 类别 | 方程模型 | 混合效应模型 | 描述 | 附注 |
|---|---|---|---|---|
| 原模型 | L1: $Y_{ij} = \beta_{0j} + r_{ij}$<br>L2: $\beta_{0j} = \gamma_{00} + u_{0j}$ | $Y_{ij} = \gamma_{00} + u_{0j} + r_{ij}$ | 单维随机效应方差分析 | 常常作为零模型以估计组间效应 |
| 随机截距模型 | L1: $Y_{ij} = \beta_{0j} + r_{ij}$<br>L2: $\beta_{0j} = \gamma_{00} + \gamma_{01} W_j + u_{0j}$<br>L1: $Y_{ij} = \beta_{0j} + \beta_{1j} X_{ij} + r_{ij}$<br>L2: $\beta_{0j} = \gamma_{00} + u_{0j}$<br>$\beta_{1j} = \gamma_{10}$ | $Y_{ij} = \gamma_{00} + \gamma_{01} W_j + u_{0j} + r_{ij}$<br>$Y_{ij} = \gamma_{00} + \gamma_{01} X_{ij} + u_{0j} + r_{ij}$ | 单维随机变量协方差分析 | 此处我们关注 L2 的预测变量 |
| 随机截距与随机斜率模型 | L1: $Y_{ij} = \beta_{0j} + \beta_{1j} X_j + r_{ij}$<br>L2: $\beta_{0j} = \gamma_{00} + u_{0j}$<br>$\beta_{1j} = \gamma_{10} + u_{ij}$ | $Y_{ij} = \gamma_{00} + \gamma_{01} X_{ij} + u_{0j} + u_{ij} X_{ij} + r_{ij}$ | 随机系数回归模型 | L1 的截距和斜率可以变化，但我们不用 L2 对其进行预测 |
| | L1: $Y_{ij} = \beta_{0j} + \beta_{1j} X_j + r_{ij}$<br>L2: $\beta_{0j} = \gamma_{00} + \gamma_{01} W_j + u_{0j}$<br>L3: $\beta_{1j} = \gamma_{10} + \gamma_{11} W_j + u_{1j}$ | $Y_{ij} = \gamma_{00} + \gamma_{01} W_j + \gamma_{10} X_{ij}$<br>$+ \gamma_{11} W_j X_{ij} + u_{0j}$<br>$+ u_{1j} X_{ij} + r_{ij}$ | 截距和斜率作为结果 | 第二层预测因素预测第一层的截距和斜率 |

## 建立与测量多层次模型

### 烟草工业数据介绍

我们使用烟草控制政策研究的一项数据来说明如何发展、检验以及解释典型多层次模型(数据可供下载,请见附录)。这项研究的主要目的是确立 1997 年至 2000 年间国会成员对烟草工业相关法案的投票行为的重要影响因素(Luke & Krauss, 2004)。因变量是投票百分比,具体而言是在 4 年间,美国某参议员或众议员“支持烟草工业”的投票占所有烟草工业相关投票的比例。例如,1998 年提出的修正法案 S1415-144 认为,应在 3 年内将每包烟的联邦烟草税提高 1.50 美元,对此,参议员 Ted Kennedy 投了否决票。根据我们的定义,此时的否决票即“支持烟草工业”投票,因为这是被烟草工业所反对的法案。这项法案最后以 40%的支持率对 58%的反对率而未通过。此处我们定义的“投票比”是先计算国会成员支持烟草工业一方的投票次数的总和,然后除以其参加的烟草相关法案投票的总次数得出。该变量的范围在 0.0(从未支持烟草工业)和 1.0(总是支持烟草工业)之间(使用百分比或比例作为模型因变量所导致的问题会在第 3 章中讨论)。

“政党”指立法者所处的党派。已有研究表明,政党在很大程度上决定着投票方向——共和党在投票中更倾向于支持烟草工业。另一个个体层面的重要变量是资金,即国会成员从烟草工业政治行动委员会(political action committees)接受资金的额度。我们假设国会议员获得的资金越多,就越有可能支持烟草工业。

除了一层变量,我们还需考虑二层单位的信息。首先,我们必须知道国会成员所代表的州,然后再搜集我们将在模型中使用的二层变量,其中最关键的便是各州烟草种植场的经济情况,通过英亩数,即 1999 年各州收获的烟草数量(以每千英亩为单位)来测量。

尽管许多软件都支持将各层信息录入在同一个数据里(如 SAS 和 S-Plus),

但 HLM 却将信息分层次整理在两个数据文件中。表 2.2 和表 2.3 即我们使用的烟草数据。一层数据文件含有将一层个案连接进入二层单位的索引信息,如该例中州的缩写。HLM 要求数据按照索引变量分类。多层次分析的最低数据要求即必须含有因变量和索引变量,而在大多数情况下,数据还会包含一系列一层和二层的预测自变量。

**表 2.2　烟草数据集一层结构**

| 姓　名 | 国会 | 州 | 投票比 | 政党 | 资金(千美元) |
|---|---|---|---|---|---|
| Murkowski | 上院 | AK | 0.84 | 共和党 | 9.2 |
| Young | 下院 | AK | 0.57 | 共和党 | 23.5 |
| Shelby | 上院 | AL | 0.64 | 共和党 | 24.2 |
| Cramer | 下院 | AL | 0.89 | 民主党 | 14.0 |
| …… | | | | | |

注:$N = 527$。

**表 2.3　烟草数据集二层结构**

| 州 | 英亩数 | 州 | 英亩数 |
|---|---|---|---|
| Ak | 0 | AR | 0 |
| AL | 0 | …… | |

注:$N = 50$。

对那些要求单个数据集的软件,数据应按照低一级的层次来组织。二层变量被分解开,保存在每一条一级单位的记录里。例如在 S-Plus 中,分析此处的烟草数据则应使用一个含有 527 条记录的数据集,每条记录针对一位国会议员。除去索引变量(州的缩写)和一层自变量,每条记录仍有"英亩数"的值,即分解后的二层变量,来自同一个州的国会议员将在这个变量上被赋予相同的值。

## 评估多层次模型是否必要

建立多层次模型的第一步就是评估是否有此必要,评估一般从以下几个方面进行:实证方面、统计方面、理论方面。下面,我们将使用烟草控制研究的例

子分别对这 3 个方面进行讨论。

图 2.2 是美国各州议员支持烟草工业的平均投票率。地图显示，各州的投票行为有着明显的不同：东南各州及平原州大多支持烟草工业利益，而新

**图 2.2 美国国会成员支持烟草工业的平均投票率**

英格兰地区则在投票中较反对这项工业。图 2.3 烟草工业政治行动委员会的资金支持与投票行为间关系的散点图更进一步显示了 5 个最大州的差别。尽管 5 个州都显示了资金支持与投票行为间的正相关，但纽约州(NY)的平均支持率却是最低的，而在加利福尼亚州(CA)和伊利诺伊州(IL)，资金支持与投票行为间的关系表现得更为强烈(更陡的斜率)。在图 2.4 中，S-Plus 或 R 软件为各州内部的线性拟合提供了更详细的线索(Becker & Cleveland, 1996)(如后文所示，S-Plus 和 R 相比其他统计软件包，可提供更丰富和灵活的图表工具)。该图同样也显示了支持性投票比例随着资金投入的增长而提高，但州际差异却更为明显。例如，俄克拉荷马州和密歇根州与大多数州不同，其资金支持和投票行为并不成正比。而且，在较小的州(议员数量较少)，此项

支持烟草工业的投票率（%）

政治行动委员会的资金支持（千美元）

**图 2.3　美国 5 个大州内资金支持与投票率的关系**

关系要弱于较大的州。

这类画图技术在搜集是否有必要使用多层次模型的实证证据时非常有用，而更正式的实证证据则是组间相关系数（ICC），它测量因变量的方差中被组别（二层单位）所解释的部分：

$$\rho = \frac{\sigma_{u_0}^2}{(\sigma_{u_0}^2 + \sigma_r^2)} \qquad [2.3]$$

此处的 $\sigma_{u_0}^2$ 和 $\sigma_r^2$ 分别是二层以及一层变量，通过用 HLM 或 SAS 中的 *Proc MIXED*，或 S-Plus 及 R 中的 *nlme* 拟合零模型得出。

表 2.4 和表 2.5 分别是使用 HLM 以及 R 拟合自由模型的部分输出结果，其中不包含任何一层或二层预测因素。

图 2.4　美国 50 个州内资金与投票率的 OLS 拟合图

$$\text{第一层：} Y_{ij} = \beta_{0j} + r_{ij} \qquad [2.4]$$
$$\text{第二层：} \beta_{0j} = \gamma_{00} + u_{0j}$$

该模型的混合效应形式则为：

$$Y_{ij} = \gamma_{00} + u_{0j} + r_{ij} \qquad [2.5]$$

上述方程中的 $Y_{ij}$ 是某个特定州内某位议员的投票比例，唯一的固定效应 $\gamma_{00}$ 是国会所有议员的平均值，而误差项则有两部分——州际差别（$u_{0j}$）和各州内议员的个体差别（$r_{ij}$）。这一零模型与方差分析中的单项随机效应模型类似。

**表 2.4 零模型的部分 HLM 输出结果**

| 第一层次随机效应 | 系数 | 可靠性估计 |
|---|---|---|
| INTRCPT1，B0 | | 0.729 |

第七次迭代的似然函数值＝ －1.559344E＋002
结果变量为 VOTEPCT
对固定效应的最终估计：

| 固定效应 | 系数 | 标准误 | *T* 比值 | 自由度 | *P* 值 |
|---|---|---|---|---|---|
| INTRCPT1，B0 | | | | | |
| INTRCPT2,G00 | 0.530942 | 0.031142 | 17.049 | 49 | 0.000 |

对方差分量的最终估计：

| 随机效应 | | 标准差 | 方差分量 | 自由度 | 卡方 | *P* 值 |
|---|---|---|---|---|---|---|
| INTRCPT1， | U0 | 0.18798 | 0.03534 | 49 | 214.13316 | 0.000 |
| Level-1， | R | 0.30436 | 0.09264 | | | |
| 协方差分量模型 | | | | | | |

注：离差＝311.868804；估计参数数目＝3。

**表 2.5 零模型的部分 R 输出结果**

最大似然估计的线性混合效应模型
数据：RSGrp

| AIC | BIC | 对数似然 |
|---|---|---|
| 319.7067 | 332.5083 | －156.8533 |

随机效应：
函数：～1|state

| | (截距) | 残差 |
|---|---|---|
| 标准差 | 0.1879767 | 0.3043614 |

固定效应：votepct～1

| | 值 | 标准误 | 自由度 | *t* 值 | *p* 值 |
|---|---|---|---|---|---|
| (截距) | 0.530942 | 0.03116992 | 477 | 17.03379 | <0.0001 |

标准化组内残差：

| 最小值 | 第一四分位 | 中值 | 第三四分位 | 最大值 |
|---|---|---|---|---|
| －2.23742030 | －0.85314980 | 0.03887117 | 0.75954570 | 2.10386071 |

样本量：527
组数：50

HLM 和 R 的输出结果都显示二层方差为 0.035，而一层方差为 0.093(R 也以标准差为单位报告随机效应，而 HLM 则同时使用标准差和方差单位)。组间相关系数为 0.035/(0.093 + 0.035) = 0.27。这表明州际差别占所有议员行为差别的 27%。较高的组间相关系数提示我们应使用多层次模型以拟合有用的各州的特征。

由于不存在一层或二层预测变量，零模型中只有一项固定效应(即 $\gamma_{00}$)被估计，其估计值为 0.53，可被解释为所有研究对象的因变量的平均值。所以，平均而言，参议员或众议员会以略高于一半的几率投票支持烟草工业。

我们使用多层模型的第二个理由是数据的结构特性。单层次一般最小二乘法模型的重要假设之一是各项观察值(以及误差项)彼此独立。但当数据出现嵌套结构时，这一假设即被违反。较高的组间相关系数同样提示我们，数据中的各项观察值之间并不彼此独立，即数据的聚类本质。相较其他州(如马萨诸塞州)，来自北加利福尼亚的议员拥有更多相同属性。政治的地方化导致各州特征如同个体特征一样影响着政治行为。

多层次模型放松独立性假设并允许相关误差结构的存在，若 OLS 被错误地运用于这类带相关误差项的聚类数据，则会低估标准误，加大犯第一类错误的几率，而多层次模型则可使估计正确无偏误。

最后且最重要的考虑因素为理论因素。一旦假设研究者使用的理论框架或建构的理论为多层次运作，则应使用多层次模型。这看似显而易见，但如第 1 章所讨论的，单层次的数据收集和分析被运用于多层次理论的情况亦相当普遍。

以烟草控制研究为例，立法员所在的州会影响其对烟草法案的态度，这是一个合理的假设。具体而言，我们会检验这一假设：相对于无烟草种植经济的州，那些烟草种植经济较强的州的立法员更有可能在投票中支持烟草工业。

我们已通过例子讨论了建立多层次模型的 3 项考虑。首先，实证考虑主要是用画图工具拟合数据，观察各州(第二层)投票行为(第一层)的显著不同。较大的组间相关系数也确认了这一点。其次，统计考虑则是意识到研究中的各项个案并不彼此独立，它们通过州而聚类，从而容易出现相关误差。最后，我们的

理论考虑是多层次模型可以测量各州的特征怎样影响投票行为。

## 从简单模型到复杂模型

建立多层次模型并不存在一种最好的办法。研究者应该考虑其研究问题是什么，其分析是解释性还是验证性的，其分析重点是参数估计、模型拟合还是预测，从而使用自己的步骤去建立模型（Harrell，2001）。但较典型的方法是由下而上地建立模型。由第一层次的自变量出发，当第一层次被满足时，考虑潜在的第二层次自变量。同时，首先考虑仅将截距作为二层自变量结果的多层次模型，当有实证或理论证据时，再引入斜率，也作为二层自变量结果的模型。我们使用烟草控制研究的数据来展示这3个步骤。当全模型建立之后，我们将讨论其他重要的统计因素，如估计技术、假设检验、效应规模、模型诊断和预测。

第一步是建立一个只含第一层预测因素的简单模型，称为“模型1”。

$$
\begin{aligned}
\text{VotePct} &= \beta_{0j} + \beta_{1j}\,(\text{Party})_{ij} + \beta_{2j}\,(\text{Money})_{ij} + r_{ij} \\
\beta_{0j} &= \gamma_{00} + u_{0j} \\
\beta_{1j} &= \gamma_{10} + u_{1j} \\
\beta_{2j} &= \gamma_{20} + u_{2j}
\end{aligned}
\qquad [2.6]
$$

此方程含有两个一层预测因素，分别为政党（Party：0＝民主党；1＝共和党）和资金（Money：从烟草工业政治行动委员会接受的资金数额，单位为千美元）。尽管我们允许一层的截距和斜率在州际变动，但此处我们并未引入二层自变量。因此，模型1有3项固定效应和4项随机效应（一项截距、两项斜率以及一项一层误差）。混合效应模型对此有更清晰的显示：

$$
\begin{aligned}
\text{VotePct} = {} & \gamma_{00} + \gamma_{10}\,(\text{Party})_{ij} + \gamma_{20}\,(\text{Money})_{ij} \\
& + u_{0j} + u_{1j}\,(\text{Party})_{ij} + u_{2j}\,(\text{Money})_{ij} + r_{ij}
\end{aligned}
\qquad [2.7]
$$

表2.6和表2.7分别是HLM和R的模型1的部分输出结果。HLM输出结果的第一部分是摘要信息，它不仅对再次分析有用，而且也可用于确保程序估计的是正确模型。

表 2.6 模型 1 的部分 HLM 输出结果

问题：ANALYSIS 2-2A-LEVEL-1 MODEL；ML

数据来源= Sagel. ssm

命令文件= whlmtemp. hlm

输出文件= G:\HLM\analyses\anl2-2a. out

二层单位最大数目= 50

迭代最大数目= 2000

估计方法:完全最大似然方法

结果变量为 VOTEPCT

固定效应模型：

| 第一层系数 | | 第二层自变量 | |
|---|---|---|---|
| INTRCPT1, | B0 | INTRCPT2, | G00 |
| PARTY slope, | B1 | INTRCPT2, | G10 |
| MONEY slope, | B2 | INTRCPT2, | G20 |

模型设定总结(函数形式)

第一层模型

Y=Bo+B1 * (PARTY)+B2 * (MONEY)+R

第二层模型

B0=G00+U0

B1=G10+U1

B2=G20+U2

结果变量为 VOTEPCT

对固定效应的最终估计(稳健标准误)

| 固定效应 | 系数 | 标准误 | *T* 比值 | 自由度 | *P* 值 |
|---|---|---|---|---|---|
| INTRCPT1, B0 | | | | | |
| INTRCPT2, G00 | 0.219654 | 0.024514 | 8.960 | 49 | 0.000 |
| PARTY slope, B1 | | | | | |
| INTRCPT2, G10 | 0.480382 | 0.021969 | 21.866 | 49 | 0.000 |
| MONEY slope, B2 | | | | | |
| INTRCPT2, G20 | 0.004650 | 0.000416 | 11.185 | 49 | 0.000 |

对方差分量的最终估计：

| 随机效应 | | 标准差 | 方差分量 | 自由度 | 卡方 | *P* 值 |
|---|---|---|---|---|---|---|
| INTRCPT1, | U0 | 0.13660 | 0.01866 | 37 | 122.02858 | 0.000 |
| PARTY slope, | U1 | 0.08967 | 0.00804 | 37 | 67.04921 | 0.002 |
| MONEY slope, | U2 | 0.00125 | 0.00000 | 37 | 36.84181 | >0.500 |
| Level-1, | R | 0.16379 | 0.02683 | | | |

协方差分量模型

注:离差=-332.009660;估计参数数目=10。

**表 2.7 模型 1 的部分 R 输出结果**

| | | | | | |
|---|---|---|---|---|---|
| 最大似然估计的线性混合效应模型 | | | | | |
| 数据 | RSGrp | | | | |
| | AIC | BIC | 对数似然 | | |
| | −308.5973 | −265.9253 | 164.2987 | | |
| 随机效应： | | | | | |
| 函数：~money + party \| state | | | | | |
| 结构：一般正定矩阵，Log-Cholesky 参数化 | | | | | |
| | 标准差 | Corr | | | |
| (截距) | 0.134575268 | (Intr) | money | | |
| money | 0.001357522 | −0.713 | | | |
| PartyRepublican | 0.086057425 | −0.813 | 0.436 | | |
| 残差 | 0.163656400 | | | | |
| 固定效应：votepct ~ money + party | | | | | |
| | 值 | 标准误 | 自由度 | $t$ 值 | $p$ 值 |
| (截距) | 0.2184723 | 0.024406748 | 475 | 8.951306 | <0.0001 |
| money | 0.0045613 | 0.000511181 | 475 | 8.923085 | <0.0001 |
| partyRepublican | 0.4824129 | 0.021664475 | 475 | 22.267461 | <0.0001 |
| 相关： | | | | | |
| | (截距) | money | | | |
| money | −0.402 | | | | |
| partyRepublican | −0.703 | −0.029 | | | |
| 标准化组内残差： | | | | | |
| | 最小值 | 第一四分位 | 中值 | 第三四分位 | 最大值 |
| | −3.62445062 | −0.59380990 | −0.01877607 | 0.54601215 | 4.46844056 |
| 样本量：527 | | | | | |
| 组数：50 | | | | | |

拟合多层次模型的关键在于估计其统计参数，即固定效应的回归参数($\gamma$)和随机效应的方差分量。应该注意，第二层误差(表中的 U0、U1、U2)本身并非统计参数，而是期望均值为 0、方差为 $\sigma_u^2$ 的潜在随机变量。

$\gamma_{00}$ 的估计值为 0.22，它并不再是投票百分比的总体均值，而是在自变量皆取值为 0 时，投票百分比的期望值。对于本数据，它意味着未从烟草工业政治行动委员会获得任何资金支持的民主党议员支持烟草工业的投票次数比例平均仅为 22%。$\gamma_{10}$ 的估计值为 0.48，这告诉我们，相较民主党，共和党支持性投票的

次数高出48%。而$\gamma_{20}=0.0046$则表示议员每接受1000美元的资金支持，其支持性投票会增长约0.46%。

对于模型的随机效应部分，我们考虑其方差分量，因而不能以“效应”解释，相反，非零的方差分量表明有未被模型化的差异存在。这一信息可帮助决定是否仍需添加其他变量到模型中的问题，或决定是否应停止增加变量。相对较大的一层(0.027)和二层(0.019)方差分量提示我们，可考虑在模型中加入更多的自变量。

## 模型估计

拟合基本多层次模型最常用的方法是各类最大似然估计法。如上述输出结果所示，一层模型是用最大似然法(ML)进行拟合的，而另一个高度相关的常用方法是约束最大似然法(REML)。那么这两种方法有何不同，而我们在使用中又该如何选择呢？实际上，这两种方法会给出完全相同的固定效应估计，而其最大的不同则在于如何计算方差分量。约束最大似然法在估计方差分量时，由于考虑固定效应的自由度，因而能得出相对于完全最大似然法偏误较小的随机效应估计。然而此差别相对较小，尤其在二层单位较多，如在超过30个的情况下，可忽略不计(Snijders & Bosker, 1999)。而最大似然法的优势则在于其离差数据可用于比较两个模型中的固定效应和随机效应(见后文的模型拟合度检验)。

表2.8分别使用最大似然法和约束最大似然法对第一层模型的随机效应进行部分估计。可以看出，由于数据中有较多的第二层单位(50个)，两种方法的结果差异非常细微，不会导致模型建构和解释的任何不同。因此在大多数情况下，我们都使用最大似然估计，除非第二层单位非常少，从而导致两种方法的结果大不相同。

模型中的方差分量皆大于0，说明仍存在相当的方差未被模型化。于是我们引入一个二层自变量，即各州收获烟草的土地面积(模型2和模型3)。首先，只考虑该变量影响第一层次截距($\beta_{0j}$)的情况：

**表 2.8 ML 与 REML 对随机效应估计的比较**

| 随机效应 | 完全最大似然法 (ML) | | | 约束最大似然法 (REML) | | |
|---|---|---|---|---|---|---|
| | 方差分量 | 卡方 | $p$ 值 | 方差分量 | 卡方 | $p$ 值 |
| 截距 | 0.01866 | 122.03 | 0.000 | 0.01942 | 122.05 | 0.000 |
| 政党斜率 | 0.00804 | 67.05 | 0.002 | 0.00860 | 66.96 | 0.002 |
| 资金斜率 | 0.00000 | 36.84 | >0.500 | 0.0000 | 36.73 | >0.500 |
| 第一层 | 0.02683 | | | 0.02686 | | |

$$
\begin{aligned}
\text{VotePct} &= \beta_{0j} + \beta_{1j}\,(\text{Party})_{ij} + \beta_{2j}\,(\text{Money})_{ij} + r_{ij} \\
\beta_{0j} &= \gamma_{00} + \gamma_{01}(\text{Acres})_j + u_{0j} \\
\beta_{1j} &= \gamma_{10} + u_{1j} \\
\beta_{2j} &= \gamma_{20} + u_{2j}
\end{aligned}
\qquad [2.8]
$$

引入截距模型(第二个方程)，我们即可测量投票行为间的差异在多大程度上可以由州内烟草经济的扩展来解释。

在逻辑上进一步扩展该模型，允许烟草种植英亩数影响第一层自变量政党和资金的斜率，则得到：

$$
\begin{aligned}
\text{VotePct} &= \beta_{0j} + \beta_{1j}\,(\text{Party})_{ij} + \beta_{2j}\,(\text{Money})_{ij} + r_{ij} \\
\beta_{0j} &= \gamma_{00} + \gamma_{01}(\text{Acres})_j + u_{0j} \\
\beta_{1j} &= \gamma_{10} + \gamma_{11}(\text{Acres})_j + u_{1j} \\
\beta_{2j} &= \gamma_{20} + \gamma_{21}(\text{Acres})_j + u_{2j}
\end{aligned}
\qquad [2.9]
$$

斜率模型(模型 3)不仅测量在一个州内烟草种植英亩数对投票行为的影响，而且还测量其与其他两个第一层自变量的交互作用。此处的参数 $\gamma_{11}$ 和 $\gamma_{21}$ 都表示层间交互作用，即第二层特征对第一层关系的影响。该模型的混合效应形式如下：

$$
\begin{aligned}
\text{VotePct} = {} & \gamma_{00} + \gamma_{01}\,(A)_j + \gamma_{10}\,(P)_{ij} + \gamma_{11}\,(A)_j\,(P)_{ij} \\
& + \gamma_{20}(M)_{ij} + \gamma_{21}(A)_j(M)_{ij} \\
& + u_{0j} + u_{1j}(P)_{ij} + u_{2j}(M)_{ij} + r_{ij}
\end{aligned}
\qquad [2.10]
$$

其中，$A$ 为英亩数，$P$ 为政党，$M$ 为资金。

**模型检验:假设检验**

表 2.9 是以上 3 个模型的重要结果。为了理解这些结果，我们需弄清如何处理个体参数估计的假设检验和模型比较。固定效应参数的显著性检验与多元回归类似。在 HLM 中，估计的参数系数除以标准误则可得到自由度为 $J-p-1$ 的 $t$ 值，其中 $J$ 代表第二层单位的个数，$p$ 代表第二层自变量的个数。大部分其他多层次软件都可用最大似然法沃德检验对固定效应参数进行显著性检验(Hox，2002)，而沃德检验可被看做标准正态分布的 $Z$ 值。

除了在 HLM 中外，沃德 $Z$ 检验也被应用于检验方差分量。而 HLM 则假设方差并非正态分布，而代之以残差的卡方检验。在解释方差分量的显著性检验时应注意几点：首先，方差以 0 为界，因而不服从正态分布。更重要的是，一个显著的方差分量，其意义并不明确。归根结底，我们期待方差并不为 0。因而，方差与效应规模类似，我们可以对之进行显著性检验或建立置信区间，但相较其显著性，关注其大小显得更有意义(这也是 R 与 S-Plus 中的 *nlme* 不报告方差分量的标准误或显著性检验的原因)。

表 2.9 说明政党和资金都是非常显著的第一层自变量 ($p<0.001$)，截距模型(模型 2)显示烟草种植英亩数同样显著，尽管其影响相对较弱。当一个州的烟草收割面积增加 1000 英亩时，支持烟草工业的投票会增加 0.05%。

斜率模型(模型 3)则展示了政党、资金和烟草经济间相对更复杂的关系。模型结果显示，烟草经济不仅显著影响州内的平均投票比例(截距效应)，同时也存在显著的层级间交互作用。交互项的负值系数表明，烟草种植的出现会降低共和党以及烟草工业资金支持对投票行为的影响。有趣的是，在引入交互作用后，截距效应从 0.0005 变至 0.0027，增长了 5 倍多。

在某些情况下，我们很难从一个复杂的多层次模型中追踪到所有效应，因而有必要基于参数估计构建简单的预测方程。例如，图 2.5 反映了在 3 个不同的烟草种植英亩数下，民主党议员的平均投票比例。底部的实线是在诸如伊利诺伊这样无烟草种植的州中，烟草工业资金支持对投票行为的影响；中部虚线

**表 2.9　3 个模型的参数估计与模型拟合比较**

| 固定效应 | 模型 1 | | | | 模型 2 | | | | 模型 3 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 系数 | 标准误 | $T$ 比值 | $p$ | 系数 | 标准误 | $T$ 比值 | $p$ | 系数 | 标准误 | $T$ 比值 | $p$ |
| 对截距的估计($\beta_{0j}$) | | | | | | | | | | | | |
| 截距($\gamma_{00}$) | 0.2196 | 0.0245 | 8.96 | 0.000 | 0.2168 | 0.0240 | 9.05 | 0.000 | 0.1828 | 0.0205 | 8.90 | 0.000 |
| 英亩数($\gamma_{01}$) | | | | | 0.0005 | 0.0001 | 3.51 | 0.001 | 0.0027 | 0.0005 | 5.10 | 0.000 |
| 对政党斜率的估计($\beta_{1j}$) | | | | | | | | | | | | |
| 政党($\gamma_{10}$) | 0.4804 | 0.0220 | 21.87 | 0.000 | 0.4793 | 0.0220 | 21.84 | 0.000 | 0.5066 | 0.0215 | 23.53 | 0.000 |
| 英亩数($\gamma_{11}$) | | | | | | | | | −0.0016 | 0.0004 | 3.60 | 0.001 |
| 对资金斜率的估计($\beta_{2j}$) | | | | | | | | | | | | |
| 资金($\gamma_{20}$) | 0.0046 | 0.0004 | 11.18 | 0.000 | 0.0045 | 0.0005 | 8.18 | 0.000 | 0.0049 | 0.0005 | 8.80 | 0.000 |
| 英亩数($\gamma_{21}$) | | | | | | | | | −0.00002 | 0.0000 | 5.50 | 0.000 |
| **随机效应** | **模型 1** | | | | **模型 2** | | | | **模型 3** | | | |
| | 标准差 | 方差分量 | 卡方 | $p$ | 标准差 | 方差分量 | 卡方 | $p$ | 标准差 | 方差分量 | 卡方 | $p$ |
| 对截距的估计($u_{0j}$) | 0.1366 | 0.0187 | 122.0 | 0.000 | 0.1310 | 0.0172 | 111.4 | 0.000 | 0.0978 | 0.0096 | 84.1 | 0.000 |
| 对政党斜率的估计($u_{1j}$) | 0.0897 | 0.0080 | 67.0 | 0.002 | 0.0900 | 0.0081 | 67.1 | 0.002 | 0.0705 | 0.0050 | 54.8 | 0.023 |
| 对资金斜率的估计($u_{2j}$) | 0.0012 | 0.0000 | 36.8 | >0.50 | 0.0014 | 0.0000 | 36.0 | >0.50 | 0.0009 | 0.0000 | 29.2 | >0.50 |
| 第一层等式($e_{ij}$) | 0.1638 | 0.0268 | | | 0.1636 | 0.0268 | | | 0.1628 | 0.0265 | | |
| **模型拟合** | 离差 | 参数 | AIC | BIC | 离差 | 参数 | AIC | BIC | 离差 | 参数 | AIC | BIC |
| | −332.0 | 10 | −312 | −269.3 | −334.5 | 11 | −312.5 | −265.5 | −353.8 | 13 | −327.8 | −272.3 |

则反映乔治亚等烟草种植情况一般(33000 英亩,1999 年)的州中的此项关系;顶部点线则是北加利福尼亚等烟草收割面积超过 200000 英亩的地区的同类关系情况。该图说明,一般而言,所获资金支持越多,民主党议员越有可能在投票中支持烟草工业。然而州内烟草种植英亩数的扩张调停了这一关系:在烟草收割面积较大的州,支持性投票率会偏高(较高截距),然而同时,资金支持对议员的影响却会减弱(较浅斜率)。

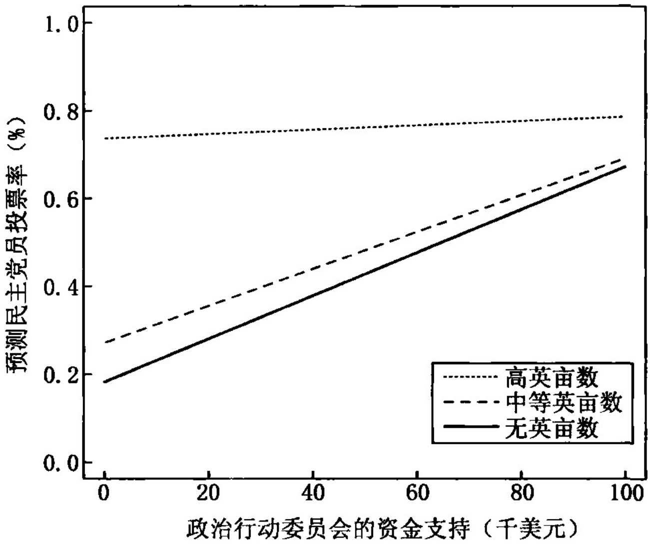



**图 2.5 根据英亩数对民主党员支持性投票率的估计**

## 模型拟合度检验——离差和判定系数

模型构建和检验的另一个重要方面是看模型与数据拟合得是否紧密。最大似然估计会产生似然值,事实上,它是多层次回归中最大似然估计被最小化的部分。若对其取自然对数并乘以－2,则得到似然值的转化形式,称为离差。离差是对数据和模型间未被拟合部分的测量,对于单个模型,它可直接解释,但在多层次模型中,它常被用于模型的比较。

模型比较适用于拟合同一数据集的两个模型,并且要求其中一个模型是另一个模型的子集(包含较少参数)。各模型离差间的差异服从卡方分布,其自由度为模型参数数量之差。

例如表 2.9 显示,模型 1 的离差为 -332,模型 2 的离差为 -334.5。较小的离差说明较好的拟合程度,而在嵌套模型中,参数较多的模型总是会有较小的离差。两个模型离差间的差异为 2.5,相对于自由度为 1(11 个参数 - 10 个参数)的卡方分布,该差异并不显著 ($p = 0.114$),这说明没有证据显示模型 2 比模型 1 提供了更好的数据拟合。然而斜率模型(模型 3)却具有显著的优越性 ($353.8 - 332 = 21.8$, $df = 3$, $p < 0.001$)。所以,若我们使用烟草收割面积作为二层自变量,则数据结果显示,它既影响第一层自变量的截距,同时又影响其斜率。

**表 2.10　用于计算 $R_1^2$ 和 $R_2^1$ 的值**

| 模　　型 | $\hat{\sigma}_r^2$ | $\hat{\sigma}_{u_0}^2$ | $H$ |
|---|---|---|---|
| 基线模型:完全无限制 | 0.093 | 0.035 | 6.23 |
| 比较模型:第二层斜率 | 0.028 | 0.005 | 6.23 |

使用离差的缺陷在于,当模型引入更多参数时,离差必然减小。为了最大程度地拟合数据,离差当然是越小越好,然而我们同时也要求模型尽量简洁,即运用最少的参数来最大化地解释数据中因变量的变动。以离差测量为基础,考虑到使用大量参数的不利因素,我们构建了另外两项模型拟合指标:赤池信息准则和贝叶斯信息准则(Akaike, 1987; Schwarz, 1978)。HLM 并不直接输出 AIC 和 BIC,但其计算却相当容易:

$$\begin{aligned} \text{AIC} &= -2LL + 2p \\ \text{BIC} &= -2LL + p\ln(N) \end{aligned} \qquad [2.11]$$

其中, $p$ 是模型中参数的总量,而 $N$ 为样本量。BIC 并不是为多层次模型而设计,故其应使用的样本规模并不明确。Singer 和 Willett(2003)认为应该使用第一层样本量,在此我们听从该建议。同离差一样,较低的 AIC 和 BIC 显示较高的模型拟合程度,而 AIC 和 BIC 的优势在于,它们可用于比较两个互不嵌套的模型。表 2.9 中计算所得的 AIC 和 BIC 证实了我们早先所做的选择,说明模型 3 是最好的模型,它有最小的 AIC 值及 BIC 值。

在一般 OLS 回归中,判定系数 $R^2$ 被用于衡量模型的拟合程度,代表因变量

的方差中为模型自变量所解释的部分。在多层次模型中，由于两方面的原因，$R^2$ 的使用更为复杂：首先，多层次模型的每一层次都有一个单独的 $R^2$；其次，若在多层次模型中使用传统方法计算 $R^2$，则有可能出现增加自变量反而导致较小的或负的 $R^2$，这显然是不合理的。Snijders 和 Bosker 的工作使我们可以相对直接地对每一层次测量 $R^2$，并使测量可得到解释。

我们并不将 $R^2$ 简单地解释为方差比例，在多层次模型中，我们将其看做预测误差的比例性减少。由于一个模型中的残差指示模型和数据之间未被拟合的部分，因而拟合得较好的模型相对比较模型而言，总是会有较小的残差。因而在二层模型中，我们有两种评价模型拟合优度的办法：首先，对第一层模型，$R_1^2$ 是预测个体结果时削减误差的比例；对第二层模型，$R_2^2$ 则是预测群体（第二层单位）均值时的削减误差比例。

根据拟合模型的输出结果，这些统计值相对易于计算。对于第一层模型，我们首先计算残差项的方差：

$$\text{var (residuals)}_1 = \sigma_r^2 + \sigma_{u_0}^2 \qquad [2.12]$$

然后，我们分别计算基线模型和比较模型的这一方差值。基线模型常常是零模型或完全自由模型，其中不包含任何第一层或第二层的自变量。而第一层预测削减误差比例则为：

$$R_1^2 = 1 - \frac{(\hat{\sigma}_r^2 + \hat{\sigma}_{u_0}^2)\text{Comparison}}{(\hat{\sigma}_r^2 + \hat{\sigma}_{u_0}^2)\text{Baseline}} \qquad [2.13]$$

若比较模型对数据拟合得较好，则第一层残差的方差较小，得出较大的 $R_1^2$。

对第二层模型的拟合优度检验遵循类似的法则。我们首先确定第二层残差的方差公式为：

$$\text{var (residuals)}_2 = \frac{\sigma_r^2}{n} + \sigma_{u_0}^2 \qquad [2.14]$$

其中，$n$ 是每个二层单位中一层单位的期望数目。分别对基准模型和比较模型计算样本的此项方差进行估计，我们可以得到第二层预测削减误差的比例：

$$R_2^2 = 1 - \frac{(\hat{\sigma}_r^2/\tilde{n} + \hat{\sigma}_{u_0}^2)\text{Comparison}}{(\hat{\sigma}_r^2/\tilde{n} + \hat{\sigma}_{u_0}^2)\text{Baseline}} \quad [2.15]$$

多层次模型的基本输出结果提供 $\hat{\sigma}_r^2$ 和 $\hat{\sigma}_{u_0}^2$ 的值。然而研究者还需提供 $\tilde{n}$ 的值，它代表任意二层单位中一层单位的典型个数。它可以来自理论，也可来自总体中样本量的期望值。在缺乏理论指引的情况下，若数据中的组规模各不相同，则可以使用第二层单位样本量的调和均值：

$$H = \frac{k}{\sum_1^k (1/n_j)} \quad [2.16]$$

其中，$k$ 是第二层单位的数量。若第二层单位的样本量并未过于不均衡，则组规模的简单均值和该调和均值会比较接近，从而可以替代调和均值。

计算 $R_1^2$ 和 $R_2^2$ 的唯一麻烦出现在使用斜率模型时。此时除截距的第二层方差 ($\sigma_{u_0}^2$) 外，每个斜率也会有方差，如 $\sigma_{u_1}^2$、$\sigma_{u_2}^2$ 等，但上述方程却只使用了截距方差。怎样处理这一问题呢？Snijders 和 Bosker(1999)建议去除随机斜率并重新拟合模型，从而仅仅生成计算所需要的两部分方差，而并不显著改变实际的参数估计。

以烟草投票行为研究为例，我们试图评估二层斜率模型(模型 3)在预测个体投票行为以及州平均投票行为时的拟合优度。表 2.10 显示了评估所需的方差分量以及调和均值(根据 Snijders 和 Bosker 的建议，我们将政党和资金作为固定效应而非随机效应重新拟合斜率模型，因而模型仅含两个方差分量：$\hat{\sigma}_r^2$ 和 $\hat{\sigma}_{u_0}^2$)。

使用本数据，$R_1^2 = 1-[(0.028+0.005)/(0.093+0.035)] = 0.742$ ①。在第二层中，$R_2^2$ 是 $1 - \frac{(0.028/6.23 + 0.005)}{(0.093/6.23 + 0.035)} = 0.81$。所以，在引入两个一层变量(政党和资金)以及一个二层变量(英亩数)后，相较零模型，比较模型的预测力被提高了大约 75%至 80%。

---

① 原书中方程错误。——译者注

## 模型评估:模型诊断

检验模型的潜在假设是否成立,这是多层次模型适合性检验的重要一环。有两项可做实证检验的重要假设:(1)第一层(组内)误差互相独立并且服从均值为 0 的正态分布;(2)随机效应服从均值为 0 的正态分布,并且在组间相互独立。这些假设可使用模型化过程中的一层和二层残差来评估。尽管此处讨论的残差分析画图技术也可由其他软件包完成,但显然 S-Plus 或 R 的 *lme* 和 *nlme* 程序更为简单和灵活。多层次模型的残差项都存储在模型拟合对象里,而这些对象可以被用于那些自动知道如何处理残差的画图程序中。

使用最终模型(表 2.9 的模型 3),我们先检验第一层残差。以各州对残差作盒形图是一种较有效的方式(图 2.6)。这类绘图法可用于检验残差是否集中

**图 2.6 模型 3 州内残差的盒形图**

于 0(竖线)及其方差是否在各组间保持不变。根据图示,尽管差异较大,但残差项确实集中于 0,而各个州之间的方差并不相等。但由于许多州的样本量都非常小,因而我们也不能过分依赖个体层面组内方差的盒形图。图 2.9 似乎显示,顶部各州的残差项倾向于为正,而底部各州残差则多为负。这仅由于州的排列顺序所致,因为 S-Plus 从因变量的最大值开始,将各州由上而下排列。

另一个常用的诊断图形是标准化残差与拟合值间的散点图,这在评估异方差问题时尤其有用。图 2.7 分两个政党显示了上述散点图,它们表示该研究中的地板效应(许多民主党人会对烟草工业投反对票,但支持性投票率不会低于 0)和天花板效应(共和党人则对烟草工业投支持票,但支持性投票率不会大于 1)。这两项效应并非致命缺陷,但在第 3 章中,我们仍将讨论如何建立避免这类问题的模型。根据图示,残差集中于 0,且并未显现出严重的异方差问题,残差

**图 2.7 模型 3 标准化残差的散点图**

变化在两个政党间基本相同。最后，我们用“来自哪个州”标识了一些异常值。这些值代表那些“反对其政党”而投票的议员，因而并不能很好地拟合模型。有趣的是，这些异常值大多来自较大的州。

最后一个可用于考虑第一层残差的图是分位比较图，又称 QQ 图(Cleveland, 1993)。该图可用于评估数据分布的正态性。若数据呈正态分布，则它们在 QQ 图中会沿一条直线排列。图 2.8 说明了我们关于烟草投票行为研究的数据，其第一层残差基本服从正态分布，仅有民主党人顶端和共和党人底端的部分极端值不服从正态分布。这些值事实上也就是图 2.7 中的异常值。

**图 2.8 模型 3 残差的 QQ 图**

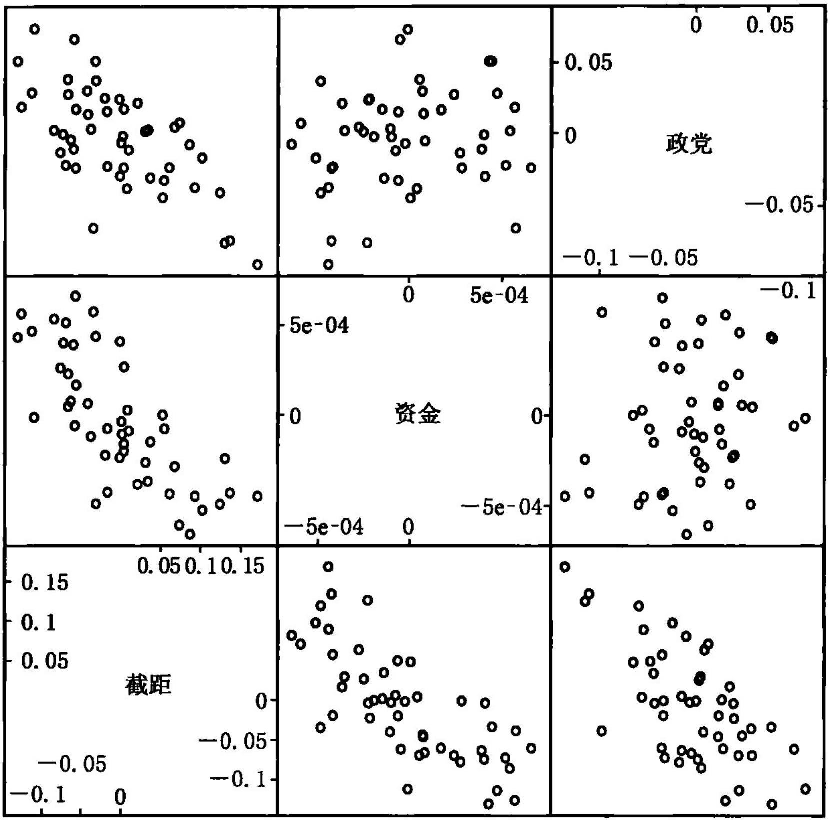
我们可以使用同类型的图形来检测第二层随机效应的假设。它们同样被假设服从于均值为 0 的正态分布。由于我们的模型有 3 项随机效应(截距、资金

斜率、政党斜率),因而我们需要逐一检验。图 2.9 是对每个随机效应分别画的一系列 QQ 图。政党效应接近于正态分布,尽管图中直线不如上图平滑,但我们应注意,相对于 527 个第一层残差,此处仅有 50 个第二层残差。

**图 2.9　模型 3 随机效应的 QQ 图**

最后,我们使用散点图矩阵来检验随机效应是否集中化,同时亦检验随机效应是否在组间彼此独立。如图 2.10 所示,这些随机效应确实集中于 0,而且并不存在严重的异方差问题。然而,在截距和资金之间以及截距和政党之间都存在着相对较大的负相关关系,而资金和政党并不彼此相关。这些关系也可能是地板效应和天花板效应的副产品。

除了上述路径之外,还存在一些读者可能感兴趣的模型诊断法。例如在多




散点矩阵

**图 2.10 模型 3 估计随机效应的散点矩阵**

层次模型中，检验第二层单位的影响也很有意义。本书不再对此进行讨论，但读者可从 Snijders 和 Bosker(1999：第 9 章)的著作中找到其步骤的简介。

## 预测：后验均值

尽管在大多数情况下，我们感兴趣的是模型的固定效应估计，但因为某些原因，我们仍需要对随机部分进行检验，尤其当我们期望得知第一层模型的截距、斜率在第二层单位间变化的具体情况时。这个过程被称做“经验贝叶斯估计”(Armitage & Colton，1998)。

试想表 2.1 顶部的完全自由模型。$\gamma_{00}$ 是决定截距的固定效应，也是 $Y_{ij}$ 的总体均值。然而，$u_{0j}$ 说明各组围绕总体均值发生变化。若我们希望对某一特

定二层单位计算 $\beta_{0j}$，则可以简单计算 $\bar{Y}_j$，即该二层单位的均值。若对该组数据本身简单进行 OLS 回归，我们同样会得到这个值。然而只有当 Y 不存在测量误差时，我们才可相信这是对特定组的估计。而且，若特定组内 Y 的测量越不可靠，我们就越倾向于用所有组的总体均值来作为估计。

这就是经验贝叶斯估计的主要原理：

$$\hat{\beta}_{0j}^{EB} = \lambda_j \hat{\beta}_{0j}^{OLS} + (1-\lambda_j)\hat{\gamma}_{00} \qquad [2.17]$$

其中，$\lambda_j$ 是第 $j$ 组中 Y 的信度。若它较高（接近 1），则截距的经验贝叶斯估计会接近简单 OLS 的估计值；而当它较低时，经验贝叶斯估计将以 Y 的总体均值生成拟合值。由于贝叶斯估计总是处在组均值和总体均值之间，而且随着信度的降低，它会偏向总体均值，所以多层次回归参数的贝叶斯估计在单组信息及整体信息之间取得了平衡。

信度的计算方程如下：

$$\lambda_j = \frac{\sigma_{u_0}^2}{(\sigma_{u_0}^2 + \sigma_r^2/n_j)} \qquad [2.18]$$

方程说明，某组的信度主要由组内第一层单位的数目（$n_j$）决定。因而样本规模较大的组，其对模型的信息贡献率要高于样本规模相对较小的组。例如，加利福尼亚州（55 位国会成员）相较俄勒冈州（仅 7 位国会成员），其投票率的测量相对可靠。

**表 2.11　美国 50 个州的经验贝叶斯估计**

| 州 | 立法员 | 截距 | 政党 | 资金 | 英亩数（千亩） |
|---|---|---|---|---|---|
| AK | 3 | 0.1937 | 0.5086 | 0.0048 | 0.0 |
| AL | 9 | 0.3660 | 0.4168 | 0.0039 | 0.0 |
| AR | 5 | 0.2824 | 0.4753 | 0.0041 | 0.0 |
| AZ | 8 | 0.1969 | 0.4986 | 0.0048 | 0.0 |
| CA | 55 | 0.1570 | 0.5646 | 0.0046 | 0.0 |
| CO | 8 | 0.2086 | 0.5370 | 0.0043 | 0.0 |
| CT | 8 | 0.0664 | 0.5201 | 0.0060 | 3.0 |
| DE | 3 | 0.1107 | 0.4915 | 0.0058 | 0.0 |
| FL | 25 | 0.2054 | 0.4755 | 0.0049 | 5.8 |
| GA | 13 | 0.2322 | 0.4891 | 0.0042 | 33.0 |
| HI | 4 | 0.1873 | 0.5042 | 0.0049 | 0.0 |

续表

| 州 | 立法员 | 截距 | 政党 | 资金 | 英亩数(千亩) |
|---|---|---|---|---|---|
| IA | 7 | 0.1241 | 0.4788 | 0.0058 | 0.0 |
| ID | 3 | 0.2249 | 0.5148 | 0.0043 | 0.0 |
| IL | 21 | 0.1901 | 0.5304 | 0.0046 | 0.0 |
| IN | 11 | 0.2731 | 0.5086 | 0.0038 | 6.5 |
| KS | 6 | 0.2518 | 0.4880 | 0.0043 | 0.0 |
| KY | 7 | 0.8259 | 0.1243 | 0.0001 | 221.6 |
| LA | 10 | 0.2943 | 0.4976 | 0.0038 | 0.0 |
| MA | 12 | 0.0595 | 0.5557 | 0.0057 | 1.3 |
| MD | 10 | 0.1177 | 0.4976 | 0.0056 | 6.5 |
| ME | 4 | 0.1135 | 0.4806 | 0.0059 | 0.0 |
| MI | 18 | 0.1542 | 0.5545 | 0.0047 | 0.0 |
| MN | 10 | 0.1933 | 0.4692 | 0.0052 | 0.0 |
| MO | 11 | 0.3270 | 0.4270 | 0.0042 | 2.3 |
| MS | 7 | 0.3157 | 0.4271 | 0.0038 | 0.0 |
| MT | 3 | 0.1526 | 0.5134 | 0.0052 | 0.0 |
| NC | 13 | 0.6767 | 0.2190 | 0.0008 | 207.8 |
| ND | 3 | 0.1440 | 0.5244 | 0.0051 | 0.0 |
| NE | 5 | 0.1680 | 0.5388 | 0.0047 | 0.0 |
| NH | 4 | 0.1898 | 0.5080 | 0.0048 | 0.0 |
| NJ | 15 | 0.1508 | 0.4354 | 0.0060 | 0.0 |
| NM | 5 | 0.1867 | 0.5385 | 0.0045 | 0.0 |
| NV | 4 | 0.1426 | 0.5247 | 0.0051 | 0.0 |
| NY | 32 | 0.0753 | 0.5879 | 0.0052 | 0.0 |
| OH | 20 | 0.1483 | 0.4849 | 0.0054 | 9.8 |
| OK | 8 | 0.2645 | 0.5226 | 0.0038 | 0.0 |
| OR | 7 | 0.1701 | 0.5270 | 0.0048 | 0.0 |
| PA | 23 | 0.1855 | 0.4767 | 0.0051 | 6.2 |
| RI | 4 | 0.0714 | 0.5349 | 0.0058 | 0.0 |
| SC | 8 | 0.3874 | 0.4420 | 0.0030 | 39.0 |
| SD | 3 | 0.1278 | 0.5262 | 0.0053 | 0.0 |
| TN | 11 | 0.3553 | 0.3736 | 0.0039 | 63.2 |
| TX | 32 | 0.3256 | 0.4367 | 0.0041 | 0.0 |
| UT | 5 | 0.1247 | 0.4953 | 0.0057 | 0.0 |
| VA | 12 | 0.3440 | 0.4013 | 0.0039 | 38.3 |
| VT | 2 | 0.1108 | 0.5055 | 0.0057 | 0.0 |
| WA | 11 | 0.1185 | 0.5369 | 0.0053 | 0.0 |
| WI | 11 | 0.1309 | 0.5792 | 0.0047 | 1.2 |
| WV | 5 | 0.2445 | 0.4729 | 0.0046 | 1.6 |
| WY | 3 | 0.2203 | 0.5140 | 0.0044 | 0.0 |

表 2.11 根据模型 3 列出了每个州的经验贝叶斯估计，各个州的议员总数以及烟草收割面积也呈现在表中。数据表明，样本量较小和烟草收割面积较少的州，其经验贝叶斯估计(EB)与拟合的 λ 值更接近。例如，将阿拉斯加州(EB一截距＝0.19，EB一政党＝0.51，EB一资金＝0.0048)和纽约州(EB一截距＝0.08，EB一政党＝0.51，EB一资金＝0.0052)的贝叶斯估计与模型 3 的一般估计(EB一截距＝0.18，EB一政党＝0.51，EB一资金＝0.0049)进行对比，则发现只拥有 3 位国会成员的阿拉斯加州，其估计的信度最低，且其贝叶斯估计与模型 3 的整体估计相当接近。而样本规模达到 32 的纽约州，其贝叶斯估计则相对远离一般模型估计。

经验贝叶斯估计常见于两类用途：首先，收缩估计可用于检验或辨识研究者感兴趣的二层单位。例如通过表 2.11，我们可以迅速确定截距最高的州(KY，即肯塔基州)，或资金与投票行为关系最显著的州(CT，即康涅狄格州和 NJ，即新泽西州)。其次，我们可以通过检验所有的经验贝叶斯预测方程来探索各州模型的差异。图 2.11 显示了每个州的预测方程，并对民主党和共和党分别作图。该图清晰地表明了政党的显著作用以及资金对支持性投票的正影响。大多数州集中在较紧密的范围内，并且民主党显示出相对较大的差异性，尤其当资金金额较大时，这种不同更加明显。少数几条异常值拟合线反映了烟草经济扩张较强的州。




**图 2.11　分政党的模型 3 的贝叶斯估计回归线**

贝叶斯估计方程还可用于考察其对原始数据的拟合程度。图 2.12 是每个州不考虑政党的收缩估计格架图，它可找出估计方程与原始数据较接近的州。一般而言，样本量较大，资金和投票行为差异性较大的州在估计和数据之间显示出了较高的一致性。例如弗吉尼亚州和俄亥俄州，其数据和贝叶斯估计间的拟合度要高于俄勒冈州（不存在接受烟草工业资金支持的国会成员）。

**图 2.12　经验贝叶斯估计的散点图**

## 中心化处理

至此，我们仍然忽略了多层次模型中的一个重要问题：第一层自变量的中心化问题。它是变量 $X$ 减去某个有意义的常数所进行的线性转换（例如，常用

的形式是减去 X 的均值)。例如，X 的总体均值中心化处理为:

$$X'_{ij}=(X_{ij}-\bar{X}_{..}) \qquad [2.19]$$

$X'_{ij}$ 即可看做偏离总体均值的离差,而不再是一个原始分数。例如,我们对资金这一变量进行总体均值中心化处理,则可将其看做相较所有国会成员接受烟草工业政治行动委员会资金支持的平均量,某一个体成员更多或较少接受资金的偏差值。若这一转化值等于 0,则说明个体接受的资金量恰好为总体均值。相较原始数值(0 表示个体从未接受此类资金),转化后的值更有意义。

事实上,中心化的一大重要作用就是提供有意义的 0 值。中心化又称“重新参数化”,常被用于一般多元回归中。然而在一般多元回归中,它并不是重要的议题,因为多元回归的关键部分(如参数估计、标准误、模型拟合等)并不随着自变量的中心化而改变。

在多层次模型中,情况则大不相同。若多层次模型有随机斜率,则第一层自变量的中心化会改变模型的某些部分(而不仅仅是转化变量的解释问题)。图 2.13 反映了该情况的原因(Hox, 2002)。在该图中, X 对 Y 的斜率在不同组之间变化,因而提示我们使用有随机斜率的多层次模型。图中标识了 $x$ 轴的两个零点。其中一个是原始的变量 X 的零点,而另一个则是 X 的

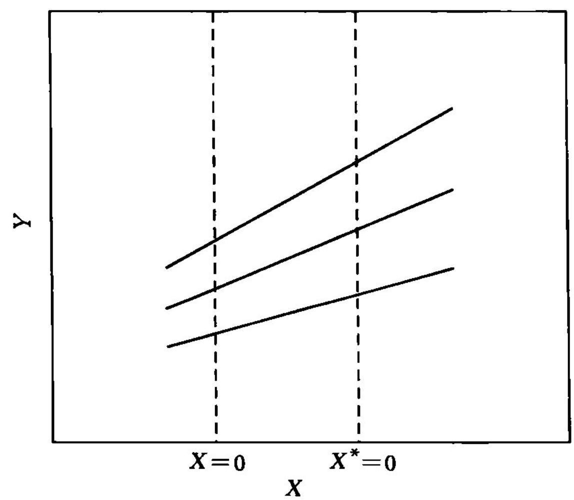



**图 2.13 中心化导致的不同截距**

一些中心化转化($X^*$)的零点。请注意,两个零点所观察到的回归线的斜率的差异不同。在 $X^*$ 的零点,斜率的差异更大,这说明第一层斜率的方差并不是常数,它会随着 $X$ 的中心化而发生变化。若我们不拟合这样一个随机斜率模型,则 3 条回归线的斜率相同,从而使得斜率的方差亦不会随中心化而改变(见图 2.1)。

表 2.12 使用烟草工业投票数据进一步说明了这一情况。表的左侧是我们之前得到的最终模型(模型 3),其中第一层自变量都未进行中心化,也就是说,资金的 0 值意味着立法员未接受政治行动委员会的任何资金支持。请注意,截距的参数系数应被解释为所有自变量为 0 时的期望值。因而在这种情况下,我们预测从未接受资金支持的民主党国会成员,其支持性投票比为 18.3%。而表的中间部分是将资金总体均值中心化后的结果,这使得截距估计变化为 0.246。由于资金变量已被中心化,因而我们的解释也会有所不同:接受资金金额为总体平均水平的民主党国会成员,其支持性投票率约为 25%。模型中仅仅与截距相关的参数估计和方差分量发生了变化,而模型的总体拟合以及资金和政党的斜率及方差分量都未受到影响。

第一层自变量的中心化还可采用另一种常见方式。除了使用总体均值外,我们还可使用组均值:

$$X'_{ij} = (X_{ij} - \overline{X}_{.j}) \qquad [2.20]$$

组均值中心化的变量可被看做在某一特定组内与平均值之间的离差,它比总体均值中心化复杂得多。不仅其解释更为困难,而且其对多层次模型的影响也更广泛。表 2.12 的右侧即组均值中心化的例子。现在,我们将截距解释为接受资金数额等同于其所在州平均值的民主党国会成员,而其支持性投票行为的比例约为 23%。与总体均值中心化不同,组均值中心化会导致整个模型的参数和方差估计都发生改变。虽然在我们的例子中并未出现较大变化,但在其他研究中,情况可能会有所不同。由于这种复杂性,所以只有在有强烈的理论动机时,才应考虑使用组均值中心化。例如,需要分辨组内回归和组间回归的情况,这常见于使用历时性数据构建增长曲线模型时(见下文)。又如,当我们对

**表 2.12　中心化与未中心化的模型比较**

| 固定效应 | 未中心化（0＝未接受任何资金） | | 总体均值中心化（0＝接受各州间平均资金值） | | 组均值中心化（0＝接受各州内平均资金值） | |
|---|---|---|---|---|---|---|
| | 系数 | 标准误 | 系数 | 标准误 | 系数 | 标准误 |
| 对截距的估计($\beta_{0j}$) | | | | | | |
| 截距($\gamma_{00}$) | 0.1828 | 0.0205 | 0.2461 | 0.0198 | 0.2305 | 0.0224 |
| 英亩数($\gamma_{01}$) | 0.0027 | 0.0005 | 0.0024 | 0.0005 | 0.0028 | 0.0005 |
| 对政党斜率的估计($\beta_{1j}$) | | | | | | |
| 政党($\gamma_{10}$) | 0.5066 | 0.0215 | 0.5066 | 0.0215 | 0.5120 | 0.0213 |
| 英亩数($\gamma_{11}$) | −0.0016 | 0.0004 | −0.0016 | 0.0004 | −0.0017 | 0.0005 |
| 对资金斜率的估计($\beta_{2j}$) | | | | | | |
| 资金($\gamma_{20}$) | 0.0049 | 0.0005 | 0.0049 | 0.0005 | 0.0044 | 0.0006 |
| 英亩数($\gamma_{21}$) | −0.00002 | 0.0000 | −0.00002 | 0.0000 | −0.00002 | 0.0000 |
| 随机效应 | 标准差 | 方差分量 | 标准差 | 方差分量 | 标准差 | 方差分量 |
| 对截距的估计($u_{0j}$) | 0.098 | 0.010 | 0.090 | 0.008 | 0.115 | 0.013 |
| 对政党斜率的估计($u_{1j}$) | 0.070 | 0.005 | 0.070 | 0.005 | 0.069 | 0.005 |
| 对资金斜率的估计($u_{2j}$) | 0.001 | 0.000 | 0.001 | 0.000 | 0.000 | 0.000 |
| 第一层等式($e_{ij}$) | 0.163 | 0.026 | 0.163 | 0.026 | 0.163 | 0.027 |
| 模型拟合 | 离差 | 参数 | 离差 | 参数 | 离差 | 参数 |
| | −353.8 | 13 | −353.8 | 13 | −335.6 | 13 |

“蛙池效应”感兴趣时，相较个体特征对结果的影响，我们更关注个体所处环境的情况，此时，组均值中心化也是有用的途径(Hox，2002)。

多层次模型中的中心化容易含混不清，而且关于各类中心化的优势也存在不少争论。关于这一议题的讨论详见《多层次模型通讯》中的一系列论文(Kreft，1995；Longford，1989；Raudenbush，1989)。下面给出一些有用的提示：

第一，根据理论基础进行中心化的选择。尽管中心化也会有一些统计上的意义，但它应该从属于研究的科学目标。

第二，若自变量的0值无意义，则应对其进行中心化，从而使其截距项可进行解释。例如，李克特式变量的取值范围在1到7之间，故我们不应使用其原始数值。因为若是如此，截距就应解释为当量表取值为0时的期望值，但整个量表却并不包含0值。

第三，二元变量和指标变量也可进行中心化。若对二元变量进行总体均值中心化，则实际上我们在解释截距时去除了该变量的影响。例如，在烟草数据中对政党和资本进行总体均值中心化，则得到的截距估计是0.51。这说明，若不考虑政党，议员投票支持烟草工业的次数一般会占一半左右。

第四，总体均值中心化仅仅影响模型中与截距相关的部分。

第五，组均值中心化在某些情况下非常有用，但其只应被用于必要的场合。

# 第3章 | 基本多层次模型的扩展

## 广义多层次模型

如前文所述，使用基本分层线性模型研究国会投票行为的一大缺陷在于，由于我们使用的因变量是一个比例值，因而违反了一般线性模型的正态分布以及同方差假设。而且，由于投票比例的分布是从0到1，因而我们使用拟合模型得到的投票行为预测值可能超出这一范围。然而，我们却很难解释为何一个国会成员会在120%的次数中对烟草工业进行支持性投票。

所幸，若对多层次模型进行扩展，则可处理许多不同类型的非连续变量和非正态变量，例如二元变量、比例变量、计数变量以及定序变量。这被称为"广义分层线性模型"(GHLM)，它会在统计模型中对因变量进行必要的转换，使之包含合适的误差分布。

以二元因变量为例来说明该问题。该变量的未转换形式取值在0到1之间，并且高度非正态分布。假设其隐含的概率分布是均值为$\mu$的二项分布，则我们用$p$来估计$\mu$值，并将其看做时间发生的概率。logit是对二项模型的典型转换：

$$\text{logit}(p) = \ln\left(\frac{p}{1-p}\right) \qquad [3.1]$$

图3.1形象地表明了logit转换的重要作用。尽管$p$是有界的，但$p$的logit转换却是无界的，而且logit($p$)的密度更接近于正态分布。

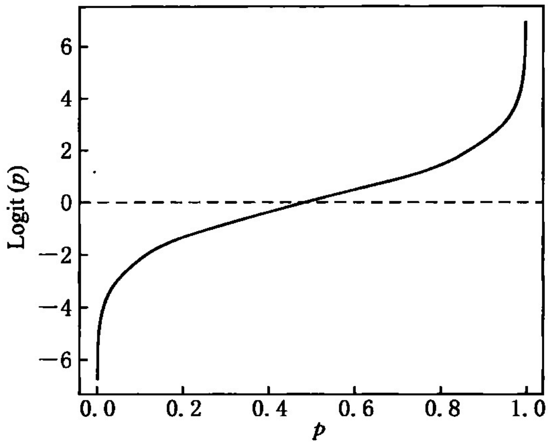



图 3.1 logit 转换图示

在广义分层线性模型中，这种转换被称做"关联函数"。我们首先使用转换式将原始的因变量 $Y$ 与新的转换变量 $\eta$ 相关联。对于二元变量，我们得到的关联函数是：

$$\eta = \text{logit}(Y) \qquad [3.2]$$

接着如第 2 章所述，我们建立一个传统的第一层预测模型：

$$\eta = \beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k \qquad [3.3]$$

模型中含有 $k$ 个自变量。请注意，此处并没有第一层的误差项。对于二元(或二项)变量，其方差由其均值决定，因而误差项不用单独估计。用于预测第一层模型中 $\beta$ 的二层模型只需如第 2 章所述建立即可。

对于不同的非正态分布数据，我们可以使用几种关联函数及其相关概率分布。表 3.1 列出了这些所谓的标准关联函数。广义分层线性模型的一大优点是可在这些关联函数与概率分布中选择适合研究数据与理论的部分使用。

以烟草工业投票行为研究为例。在第 2 章中，我们建立模型估计对烟草工业的支持性投票在总投票次数中所占的比例。然而，该比例实际上却是综合国会成员在第 105 届和第 106 届国会(从 1997 年到 2000 年)中，由一系列的个人投票行为得来的。因此，为了替代这一综合比例信息，我们使用广义分层线性

表 3.1 GHLM 的关联函数

| 因变量 | 关联函数 | 公　式 | 概率分布 |
|---|---|---|---|
| 二分变量 | logit | $\eta=\ln\left(\frac{Y}{1-Y}\right)$ | 二项分布 |
| 比例变量 | logit | $\eta=\ln\left(\frac{Y}{1-Y}\right)$ | 二项分布 |
| 频数变量 | log | $\eta=\ln(Y)$ | 泊松分布 |
| 次序变量 | 累积 logit | 见 Raudenbush 和 Bryk(2002:319) | 多项分布 |

注:引自 Raudenbush 和 Bryk(2002:333)。

模型来预测对任一法案或修正案的支持性投票的可能性。

在这个模型中,因变量投票是二元变量,若国会成员投票反对烟草工业,则编码为 0,支持烟草工业则编码为 1。我们仍试图建立多层次模型,但此时模型层次发生了改变。在模型第一层,我们测量和预测对单个法案的投票情况。而这些投票情况嵌套于每位国会成员的个人行为之中,所以此时第二层单位是个人。我们试图对某法案的支持性投票概率进行模型化,则正式模型为:

$$\begin{aligned}\text{第一层}:\eta_{ij}&=\text{logit}(Y_{ij})\\ \eta_{ij}&=\pi_{0j}\\ \text{第二层}:\pi_{0j}&=\beta_{00}+\beta_{01}\text{Party}_j+\beta_{02}\text{Money}_j+u_{0j}\end{aligned} \qquad [3.4]$$

模型中使用了 logit 关联函数,而 $Y_{ij}$ 是第 $j$ 位国会成员对第 $i$ 部法案的投票情况。模型中仅含有第二层的个体自变量,包括政党和接受的资金支持金额。如果我们有相关的法案层次的自变量(例如,该投票是针对全法案还是针对修正案,或该法案的倡议者的政党),则可进一步将其加入第一层模型中。但在本例中,我们仅估计第一层截距,即国会成员对烟草工业进行支持性投票的平均概率。请再次注意,第一层模型中不含有误差项。该模型同样仅含有一项随机效应,即个体间投票行为的差异($u_{0j}$)。

表 3.2 是二元投票行为数据的广义分层线性模型的拟合结果。与之前的结果一样,所有的系数都显著。然而,不同于之前可以直接解释的模型结果,现在得到的系数需要被转换成原始形式再进行解释。此处需要使用关联函数的反

函数，如 logit 函数的反函数为 logistic 函数：

$$Y = \text{logistic}(\beta_0 + \beta_1 X_1 + \beta_2 X_2) = \frac{e^{(\beta_0+\beta_1 X_1+\beta_2 X_2)}}{1 + e^{(\beta_0+\beta_1 X_1+\beta_2 X_2)}} \qquad [3.5]$$

**表 3.2 二层次 GHLM 模型结果**

| 固定效应 | 系数 | 标准误 | $T$ 比值 | $p$ |
|---|---|---|---|---|
| 对截距的估计($\pi_{0j}$) | | | | |
| 截距($\beta_{00}$) | −1.721 | 0.0763 | 22.54 | 0.000 |
| 政党($\beta_{01}$) | 2.544 | 0.0987 | 25.78 | 0.000 |
| 资金($\beta_{02}$) | 0.0325 | 0.0029 | 11.36 | 0.000 |
| 随机效应 | 标准差 | 方差分量 | 卡方 | $p$ |
| 截距($u_{0j}$) | 0.834 | 0.696 | 1439.0 | 0.000 |
| 分散指数 | 0.937 | | | |
| 模型拟合 | 离差 | 参数 | AIC | BIC |
| | 7636.8 | 4 | 7644.8 | 7672.5 |

因而，为了知道从未接受资金的民主党国会成员对烟草工业进行支持性投票的预测概率（政党＝0，资金＝0），我们需要计算 logistic(−1.721) = $[(e^{-1.721})/(1+e^{-1.721})] = 0.152$；而对未接受资金支持的共和党国会议员（政党＝1，资金＝0），我们应计算 logistic(−1.721＋2.544)，得出的预测概率为 0.695；而对接受资金 10000 美元的共和党国会成员（政党＝1，资金＝10），logistic(−1.721＋2.544＋10 * 0.0325) = 0.759。这些结果与表 2.9 中模型 3 的结果类似。

绘制预测图会使广义分层线性模型的解释更为容易。图 3.2 分别显示了民主党议员和共和党议员的支持性投票概率预测。该图也强调了模型非线性的本质。此外，使用广义分层线性模型使我们的预测有意义，并且不再超出概率的取值范围。

前文已经指出，二项模型在第一层模型中不含单独的误差—方差项，因为方差是总体均值的函数。然而，有些软件，如 HLM 或 R、S-Plus 却会计算第一层的误差方差的量化因子，用以测量观察误差在多大程度上服从理论二项误差

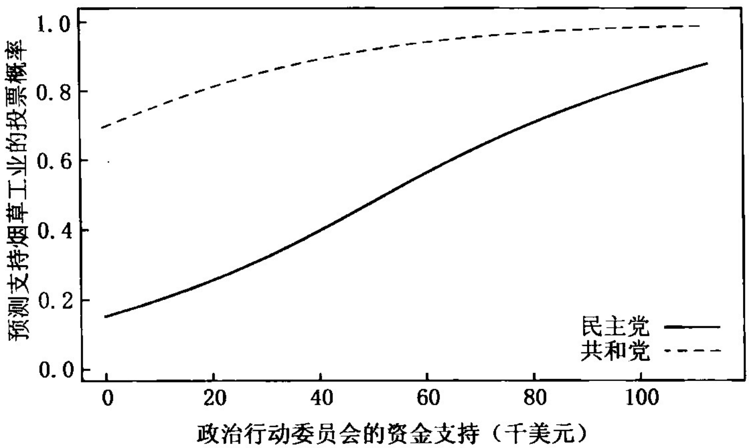



**图 3.2　二层次 GHLM 模型关于投票概率的预测结果**

分布。若量化因子为 1,则说明观察误差与理论分布完美拟合;若数值小于 1,则称做“低度分散”,大于 1 则为“过度分散”。过度分散和低度分散表明模型可能存在设定误差、大影响的异常值,或是缺少某个关键层级。分散指数是对二元/二项模型的一个很好的诊断,若出现较大或较小的值,则应对模型进行严格的检验。表 3.2 显示分散指数为 0.94,属于轻微的低分散情况,可能不足以造成较大问题。

我们应该注意广义分层线性模型中的离差、赤池信息准则、贝叶斯信息准则的解释问题,因为它们不适合用最大似然法和约束最大似然法进行估计。相反,许多软件包使用惩罚拟似然估计(PQL)。它输出似然值的渐进近似,因而离差以及在离差基础上获得的信息统计值(如赤池信息)都不再可靠,尤其当样本量很小时。其技术细节和其他非线性估计技术在 Raudenbush 和 Bryk(2002)以及 Pinheiro 和 Bates(2002)的著作中可以找到。

## 三层模型

之前我们考虑的仅是二层模型。首先在第 2 章中,我们认为个体层面的投票行为受到州一级的因素影响。接着在第 3 章中,我们对单个法案的投票结果

进行模型化，并认为其受到个体层次特征的影响。如前文所讨论的，使用广义分层线性模型拟合二元投票行为具有相当的统计优势。然而，仅仅使用二层模型将投票行为看做个体属性的函数却不允许我们测量这些个体属性（如政党和资金）在州一级的随机效应。

所幸分层线性模型和广义分层线性模型可以处理二层以上的更多层次。在社会和科学研究中，三层模型并非罕见。例如，教育的数据常在3个层次上收集：学生嵌套于班级，而班级又嵌套于学校。构建三层模型所需考虑的问题基本与二层模型相同，但应特别注意清楚地定义和合理地测量各层的假设效应。

尽管从二层模型到三层模型的统计扩展看来相对直接，但要清楚地说明随机效应却并不十分容易。在二层模型中，一层截距和斜率在第二层中是随机的。但在三层模型中，其一层截距和斜率在第二层和第三层中都是随机的（此处的随机意味着第一层截距和斜率在高层的各单位间变动）。假设我们对不同学校、不同班级的阅读能力感兴趣，并假设平均阅读成绩（第一层截距）和社会经济地位指数（SES）对阅读能力的影响（第一层斜率）在班级（第二层）和学校（第三层）之间各不相同。

此外，二层模型中的自变量也可在第三层单位变动。若我们在第二层中加入教师经验，且假设其影响在学校间不同，则在第三层模型中，应将其作为随机效应引入。对于此例，方程3.6是可能的分层线性模型：

$$
\begin{aligned}
\text{第一层：} Y_{ijk} &= \pi_{0jk} + \pi_{1jk}(\text{SES})_{ijk} + \varepsilon_{ijk} \\
\text{第二层：} \pi_{0jk} &= \beta_{00k} + \beta_{01k}(\text{Exp})_{jk} + r_{0jk} \\
\pi_{1jk} &= \beta_{10k} + \beta_{11k}(\text{Exp})_{jk} + r_{1jk} \\
\text{第三层：} \beta_{00k} &= \gamma_{000} + u_{00k} \\
\beta_{01k} &= \gamma_{010} + u_{01k} \\
\beta_{10k} &= \gamma_{100} + u_{10k} \\
\beta_{11k} &= \gamma_{110} + u_{11k}
\end{aligned}
\qquad [3.6]
$$

此处的$Y_{ijk}$是第$k$所学校第$j$个班级中第$i$名学生的阅读成绩。社会经济地位指数（SES）是第一层自变量，教师经验（Exp）是第二层自变量。所有第一层、第

二层的截距和斜率皆被处理为随机效应。这组方程可转化为混合模型方程：

$$\begin{aligned} Y_{ijk} = &\gamma_{000} + \gamma_{010}(\text{Exp}) + \gamma_{100}(\text{SES}) + \gamma_{110}(\text{Exp})(\text{SES}) \\ &+ u_{00k} + u_{01k}(\text{Exp}) + u_{10k}(\text{SES}) + u_{11k}(\text{Exp})(\text{SES}) \\ &+ r_{0jk} + r_{1jk}(\text{SES}) + \varepsilon_{ijk} \end{aligned} \qquad [3.7]$$

方程 3.6 清楚地表明了模型的层次，并且显示第一层和第二层各有一个自变量。而方程 3.7 则使模型的 4 项固定效应与 7 项方差分量(随机效应)更清晰。尤其是固定效应的层间交互作用($\gamma_{110}$)，在方程 3.6 中很容易被忽略。

广义分层线性模型同样可以被扩展至 3 层。我们的投票行为数据就为三层二元模型提供了很好的例子。之前我们曾对数据做了两个不同的二层拟合：一个是将国会成员嵌套于州内，另一个则是将对法案的个人投票嵌套于国会的每一个成员中。结合这两项考虑，我们可以构建一个三层模型，其中个人投票嵌套于国会成员中，而国会成员又嵌套于州内。

方程 3.8 是这一统计模型的正式形式。我们仍然使用带 logit 关联函数的二项模型。政党和资金支持是国会成员层次(第二层)的自变量，英亩数是第三层次的自变量，而且我们假设其只影响二层截距，并不影响二层斜率。

$$\begin{aligned} \text{第一层：} \eta_{ijk} &= \text{logit}(Y_{ijk}) \\ \eta_{ijk} &= \pi_{0jk} \\ \text{第二层：} \pi_{0jk} &= \beta_{00k} + \beta_{01k}(\text{Party})_{jk} \\ &\quad + \beta_{02k}(\text{Money})_{jk} + r_{0jk} \\ \text{第三层：} \beta_{00k} &= \gamma_{000} + \gamma_{001}(\text{Acres})_k + u_{00k} \\ \beta_{01k} &= \gamma_{010} + u_{01k} \\ \beta_{02k} &= \gamma_{020} + u_{02k} \end{aligned} \qquad [3.8]$$

方程 3.9 是单个混合模型方程，它表明模型有 4 项混合效应和 4 项随机效应。由于英亩数只被允许影响截距，因而不存在层间交互效应。

$$Y_{ijk} = \text{logistic}\left\{\begin{aligned} &\gamma_{000} + \gamma_{001}(\text{Acres})_k + \gamma_{010}(\text{Party})_{jk} + \gamma_{020}(\text{Money})_{jk} \\ &u_{00k} + u_{01k}(\text{Party})_{jk} + u_{02k}(\text{Money})_{jk} + r_{0jk} \end{aligned}\right\} \qquad [3.9]$$

表 3.3 是对该模型的拟合结果，它们与之前模型的结果非常一致。政党、资金和烟草收割面积都是投票行为的显著影响变量。三层模型中，资金的效应略小于二层模型，这可能是由于资金的部分效应被州内的烟草经济解释了——那些所在州内存在烟草工业的议员相较其他不存在烟草工业的州的议员，所接受的资金金额较高。图 3.3 是拟合模型的预测图，其中包含 3 个级别的烟草经济。图形的大致形状（政党和资金的影响）与二层模型类似。该图说明，尽管烟草经济在模型中是显著变量，但除非烟草收割面积较大，例如达到 100000 英亩，否则其作用微乎其微。因而对大多数州而言，只有政党和资金支持的作用是值得考虑的。

**表 3.3 三层 GHLM 模型结果**

| 固定效应 | 系数 | 标准误 | $T$ 比值 | $p$ |
|---|---|---|---|---|
| 对截距的估计($\pi_{0jk}$) | | | | |
| 对截距的估计($\beta_{00k}$) | | | | |
| 截距($\gamma_{000}$) | −1.644 | 0.1413 | 11.62 | 0.000 |
| 英亩数($\gamma_{001}$) | 0.0060 | 0.0020 | 2.96 | 0.005 |
| 政党($\gamma_{010}$) | 2.450 | 0.1144 | 21.41 | 0.000 |
| 资金($\gamma_{020}$) | 0.0249 | 0.0035 | 7.04 | 0.000 |
| 随机效应 | 标准差 | 方差分量 | 卡方 | $P$ |
| 截距—1($r_{0j}$) | 0.647 | 0.419 | 781.7 | 0.000 |
| 截距—2($u_{00}$) | 0.770 | 0.593 | 89.7 | 0.000 |
| 斜率—政党($u_{01}$) | 0.397 | 0.158 | 47.4 | 0.117 |
| 斜率—资金($u_{02}$) | 0.011 | 0.000 | 37.5 | 0.448 |
| 分散指数 | 0.943 | | | |
| 模型拟合 | 离差 | 参数 | AIC | BIC |
| | 7615.8 | 5 | 7625.8 | 7660.4 |

拟合模型的另一个重要部分是随机效应。在二层模型中，个人层次模型的标准差是 0.834，而在三层模型中，个人层次差异相对较低(0.647)，事实上小于组间差(0.770)。这说明州与州之间投票行为的差异大于州内部议员投票行为间的差异。而两个个人层次斜率的随机效应都小于截距，正支持了我们不将英亩数作为斜率的自变量的假设。

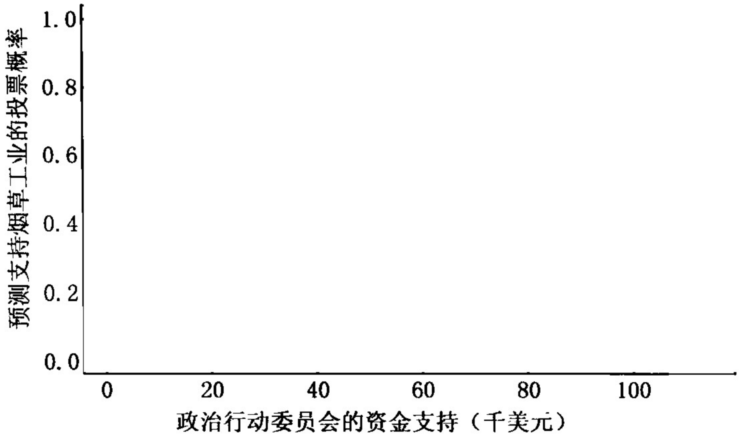



**图 3.3　三层 GHLM 模型对投票概率的预测**

## 分层纵向数据：嵌套于个体的时点

在使用多层次模型时，常见的是个体对象嵌套在一个物理或社会环境中的情况，例如街区中的个人或健康维护组织中的医院（见表 1.2）。然而，如同个体对法案的投票行为研究一样，多层次模型也可用于多项观察嵌套于单一对象的情况。尤其当我们的研究兴趣为历时性变化的结构和影响因素时，多层次模型可用于分析纵向数据。

相对于传统分析方法，多层次模型在研究纵向数据时具有诸多优势。图 3.4 是典型历时性研究中可能经常出现的散乱数据。第 001 号对象代表一种理想状况，即参与者在第一个月接受初访，而在其后每 6 个月进行一次追访，共进行 4 次追访。第 002 号则是放弃调查或在追访时遗失的例子。第 003 号是在某几个时间点上出现了缺失值。第 004 号则是仅在两个时间点具有有效数据。第 005 号虽然有初访和 3 次追访的完整数据①，但其访问发生的时间与计划不符。最后，第 006 号不仅有缺失值，而且其时间点也与计划不符。

① 此处原文有误。

| 对象 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 001 | * | | | | | | * | | | | | | * | | | | | | * | | | | | | * |
| 002 | * | | | | | | * | | | | | | * | | | | | | | | | | | | |
| 003 | * | | | | | | * | | | | | | | | | | | | | | | | | | * |
| 004 | * | | | | | | | | | | | | | * | | | | | | | | | | | |
| 005 | * | | | | | * | | | | | | | | | | | | | | | * | | | * | |
| 006 | | | | | | * | | | | | | | | | | | | | | | | | | * | |

图 3.4 一个较混乱的面板数据集

许多传统历时性研究，如重复测量的多元方差分析，都不能较容易地处理不均衡或包含缺失值的情况，或时点不均匀的纵向数据。而多层次模型则相对有效且灵活，它可使用任何能够获得的数据，并且可以拟合在不同时点上采集的数据，对历时变化的模式进行预测。

为了举例说明纵向数据的多层次模型研究，我们使用全国卫生统计中心与全国老年研究所在 1984 年到 1990 年间合作进行的两年一度的纵向调查数据，该调查的研究对象是 70 岁或以上的老年人，被称做“老年人纵向数据”(LSOA)。本书使用的数据提取自 ICPSR 上公开发表的数据。该数据的重要目的之一是测量老年人机能状态和生活安排的改变，使之特别适合多层次模型。参与者至少为 70 岁，1984 年，他们被纳入调查并接受了第一次访问，在之后直至 1990 年，他们每两年接受一次再访问。

表 3.4 和表 3.5 是数据结构和研究中使用变量的情况。第一张表是访谈层次(第一层)的数据集。样本序列号(Case ID)是参与者的号码，并且关联着两个数据集。困难活动数量(NumADL)是受访者在日常生活中面临困难的活动数量，该变量的取值从 0 至 7，其覆盖的困难包括洗浴、穿衣、进食、上下床及上下轮椅、行走、外出和如厕。因而该变量可看做对机能状态的测量，较高的活动数量表明更差的机能状态。婚姻状况(Married)是记录参与者在受访时是否已婚的二元变量。由于婚姻状态是可变的，因而它被作为时变变量放在访谈层次的数据集内，而不是放在个体层次的数据集中。访谈序列(Wave)是时间变量，表

明此次访谈是第几次，它将被用于拟合线性变化。它是已经被减去 1 而重新参数化后的变量，所以对于第一次访谈而言，序列＝0。序列平方（$Wave^2$）即访谈序列的平方，将被用于检验二次变化。时点 1 至时点 4（IND1—IND4）是 HLM 软件用于指定访谈时点的变量（其他软件，如 SAS *PROC Mixed* 或 S-Plus *nlme*，并不需用户创造这类变量）。

**表 3.4 LSOA 一层数据**

| 样本序列号 | 困难活动数量 | 婚姻状况 | 访谈序列 | 访谈序列$^2$ | 时点 1 | 时点 2 | 时点 3 | 时点 4 |
|---|---|---|---|---|---|---|---|---|
| 1 | 0 | 否 | 0.00 | 0.00 | 1.00 | 0.00 | 0.00 | 0.00 |
| 2 | 0 | 否 | 0.00 | 0.00 | 1.00 | 0.00 | 0.00 | 0.00 |
| 2 | 0 | 否 | 1.00 | 1.00 | 0.00 | 1.00 | 0.00 | 0.00 |
| 2 | 0 | 否 | 2.00 | 4.00 | 0.00 | 0.00 | 1.00 | 0.00 |
| 2 | 1 | 否 | 3.00 | 9.00 | 0.00 | 0.00 | 0.00 | 1.00 |
| 3 | 0 | 否 | 0.00 | 0.00 | 1.00 | 0.00 | 0.00 | 0.00 |
| 3 | 0 | 否 | 3.00 | 9.00 | 0.00 | 0.00 | 0.00 | 1.00 |
| 4 | 0 | 否 | 0.00 | 0.00 | 1.00 | 0.00 | 0.00 | 0.00 |
| 4 | 5 | 否 | 1.00 | 1.00 | 0.00 | 1.00 | 0.00 | 0.00 |
| 4 | 0 | 否 | 2.00 | 4.00 | 0.00 | 0.00 | 1.00 | 0.00 |
| 4 | 1 | 否 | 3.00 | 9.00 | 0.00 | 0.00 | 0.00 | 1.00 |

**表 3.5 LSOA 二层数据**

| 样本序列号 | 性别 | 开始年龄 |
|---|---|---|
| 1 | 女性 | 70 |
| 2 | 男性 | 87 |
| 3 | 女性 | 71 |
| 4 | 男性 | 78 |

参与者层次的数据文件较为简单，表 3.5 即是。开始年龄（Age84）是研究开始时参与者的年龄；性别（Male）则是指示参与者性别的二元变量，其中男性编码为 1，女性编码为 0。只有在所有时间皆不发生改变的变量才被放入这一个体层次数据集。

我们的基本目标是模型化 6 年间日常生活出现问题的轨迹。同时，我们关

注年龄、性别、婚姻状况如何影响这一轨迹。数据包含 7417 位参与者的 20283 次访谈信息，平均每人达 2.73 次。由于每个参与者应被调查 4 次，所以大约有 32%的缺失值。事实上，仅有 2330 位参与者(31.4%)完成了 4 次访谈。换言之，使用重复测量的方差分析来研究这一数据并不可取，而多层次模型则可以在不损失样本与信息的情况下分析整个数据。

表 3.6 是每次调查计算的活动数量均值。基于样本均值，我们可以观察到机能的逐年下降，因而我们首先要拟合的模型是机能的线性和二次变化以及年龄和性别的影响：

**表 3.6 不同时点上的日常生活平均困难数目**

| | 基准年 | 2 年后 | 4 年后 | 6 年后 |
|---|---|---|---|---|
| 平均困难数目 | 0.72 | 1.21 | 1.19 | 1.32 |

$$
\begin{aligned}
&\text{第一层：} Y_{ij} = \beta_{0j} + \beta_{1j}(\text{Wave})_{ij} + \beta_{2j}(\text{Wave}^2)_{ij} + r_{ij} \\
&\text{第二层：} \beta_{0j} = \gamma_{00} + \gamma_{01}(\text{Male})_j + \gamma_{02}(\text{Age} - \overline{\text{age}})_j + u_{0j} \\
&\qquad\quad \beta_{1j} = \gamma_{10} + u_{1j} \\
&\qquad\quad \beta_{2j} = \gamma_{20} + u_{2j}
\end{aligned}
\qquad [3.10]
$$

请注意，我们允许时间斜率在不同个体间变化，但并不试图将其与个人层次的变量相关联。同时，我们对个人层次的变量年龄进行中心化，取其与总体均值的差。若非如此，我们无法解释截距项，因为它反映了 0 岁时的个人情况。

该模型(模型 1)的结果显示在表 3.7 的左侧。图 3.5 是根据这些结果所绘制的预测图，该图仅针对 1984 年调查时，年龄等于总体均值的个体。男性和女性在调查开始时的平均困难活动数量都小于 1，并且在其后的 6 年间稳定增长。访谈序列的参数估计为 0.473，表明每两年困难活动数量约增长 0.5。尽管序列平方在模型中也显著，但其系数仅为−0.046，远小于线性效应。二次效应可以简单解释为变化率，因而该结果说明随着时间的推移，机能水平下降的速率在降低。男性机能略好于女性(−0.210)，并且 1 岁的年龄增长会导致困难活动数量增多 0.073。

预测困难活动数量

研究开始后的年份

**图 3.5 生活困难随时点变化的预测射线**

随机效应的方差估计显示，个体间方差(1.070)和个体内方差(1.014)大致相等。考虑到固定效应的大小，说明还有相当部分的方差未被模型拟合。

表 3.7 中的模型 2 和模型 3 是初始模型的补充。模型 2 加入第一层次的时变变量婚姻状况(方程 3.11)，检验已婚个体较之未婚个体，其机能状况较好的假设。婚姻状况是作为固定效应进入模型的，否则 HLM 将无法集合，这可能是由于个体间及个体内部婚姻状况的差异较小造成的。结果显示，婚姻状况并不是影响机能状态的显著变量。

$$\text{第一层：} Y_{ij} = \beta_{0j} + \beta_{1j}(\text{Wave})_{ij} + \beta_{2j}(\text{Wave}^2)_{ij} + \beta_{3j}(\text{Married})_{ij} + r_{ij}$$

$$\text{第二层：} \beta_{0j} = \gamma_{00} + \gamma_{01}(\text{Male})_j + \gamma_{02}(\text{Age} - \overline{\text{age}})_j + u_{0j}$$

$$\beta_{1j} = \gamma_{10} + u_{1j}$$

$$\beta_{2j} = \gamma_{20} + u_{2j} \qquad [3.11]$$

$$\beta_{3j} = \gamma_{30}$$

模型 1 仅仅将性别作为重要效应加以考虑。而我们可以通过加入层间交互项来检验男性和女性是否具有同样的机能轨迹。模型 3 就引入了这些交互项(方程 3.12)，结果显示，性别和线性效应以及性别和二次效应之间的交互作用皆不显著。

**表 3.7　LSOA 纵向数据的参数估计和模型拟合**

| 固定效应 | 模型 1 | | | | 模型 2 | | | | 模型 3 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 系数 | 标准误 | $T$ 比值 | $p$ | 系数 | 标准误 | $T$ 比值 | $p$ | 系数 | 标准误 | $T$ 比值 | $p$ |
| 对截距的估计($\beta_{0j}$) | | | | | | | | | | | | |
| 截距($\gamma_{00}$) | 0.802 | 0.021 | 37.4 | 0.000 | 0.818 | 0.024 | 34.3 | 0.000 | 0.816 | 0.024 | 34.0 | 0.000 |
| 性别($\gamma_{01}$) | −0.210 | 0.034 | 6.1 | 0.000 | −0.189 | 0.037 | 5.1 | 0.000 | −0.185 | 0.038 | 4.87 | 0.000 |
| 年龄($\gamma_{02}$) | 0.073 | 0.003 | 24.5 | 0.000 | 0.072 | 0.003 | 23.6 | 0.000 | 0.003 | 0.003 | 23.6 | 0.000 |
| 对访谈序列的斜率的估计($\beta_{1j}$) | | | | | | | | | | | | |
| 访谈序列($\gamma_{10}$) | 0.473 | 0.025 | 19.0 | 0.000 | 0.472 | 0.025 | 18.9 | 0.000 | 0.470 | 0.031 | 15.1 | 0.000 |
| 性别($\gamma_{11}$) | | | | | | | | | 0.006 | 0.052 | 0.1 | 0.900 |
| 对序列平方的斜率的估计($\beta_{2j}$) | | | | | | | | | | | | |
| 序列平方($\gamma_{20}$) | −0.046 | 0.008 | 5.4 | 0.000 | −0.046 | 0.008 | 5.4 | 0.000 | −0.043 | 0.010 | 4.13 | 0.000 |
| 性别($\gamma_{21}$) | | | | | | | | | −0.007 | 0.018 | 0.4 | 0.679 |
| 对婚姻状况的斜率的估计($\beta_{3j}$) | | | | | | | | | | | | |
| 婚姻状况($\gamma_{30}$) | | | | | −0.051 | 0.033 | 1.5 | 0.128 | −0.051 | 0.033 | 1.5 | 0.130 |
| **随机效应** | 标准差 | | 方差分量 | | 标准差 | | 方差分量 | | 标准差 | | 方差分量 | |
| 截距($u_{0j}$) | 1.070 | | 1.144 | | 1.070 | | 1.144 | | 1.070 | | 1.144 | |
| 访谈序列斜率($u_{1j}$) | 0.704 | | 0.495 | | 0.703 | | 0.494 | | 0.703 | | 0.494 | |
| 序列平方斜率($u_{2j}$) | 0.195 | | 0.038 | | 0.195 | | 0.038 | | 0.195 | | 0.038 | |
| 第一层截距($r_{ij}$) | 1.014 | | 1.029 | | 1.014 | | 1.029 | | 1.014 | | 1.029 | |
| **模型拟合** | 离差 | 参数 | AIC | BIC | 离差 | 参数 | AIC | BIC | 离差 | 参数 | AIC | BIC |
| | 73967 | 12 | 73991 | 74074 | 73965 | 13 | 73991 | 74081 | 73964 | 15 | 73994 | 74098 |

$$
\begin{aligned}
&\text{第一层}: Y_{ij} = \beta_{0j} + \beta_{1j}(\text{Wave})_{ij} + \beta_{2j}(\text{Wave}^2)_{ij} + r_{ij} \\
&\text{第二层}: \beta_{0j} = \gamma_{00} + \gamma_{01}(\text{Male})_j + \gamma_{02}(\text{Age} - \overline{\text{age}})_j + u_{0j} \\
&\qquad\quad \beta_{1j} = \gamma_{10} + \gamma_{11}(\text{Male})_j + u_{1j} \\
&\qquad\quad \beta_{2j} = \gamma_{20} + \gamma_{21}(\text{Male})_j + u_{2j} \\
&\qquad\quad \beta_{3j} = \gamma_{30}
\end{aligned}
\qquad [3.12]
$$

统计系数表明，我们并没有理由拒绝模型 1 而选择模型 2 或模型 3。在模型拟合的 3 个数据中，贝叶斯信息准则最能区分最优模型。由于贝叶斯信息在选择最小离差的基础上考虑样本量的影响，所以其优势在处理大样本数据时更明显。由于模型 1 具有最小的贝叶斯信息值，因而模型 1 即最优模型。

此处纵向数据的模型建构步骤大致类似我们之前用到的多层次模型。然而，处理纵向数据还需考虑另一个重要问题。在非纵向数据中，我们总是假设误差的正态分布以及独立性，但独立性假设却常常不适用于纵向数据。因此，研究者需为纵向数据选择一个合适的误差替代结构。大部分软件可以模拟这一结构，虽然为了使用者的方便，这些结构的设置各不相同。

表 3.8 就 3 种不同的协方差结构向我们展示了模型 1 的拟合优度指标。最一般的误差结构是无限制或非结构化的，它对误差项不进行任何假设，因而允许任何形式的相关性误差存在。无限制的协方差结构总是会得出最大的随机参数值，这是因为每一段时间间隔都会由数据估计出独立的协方差。所以，尽管无限制误差模型的离差总是最小，但它却是最不简约的模型。同时，对误差结构进行一定限制的模型不仅在理论上更为合理，而且也更易于估计。综上所述，无限制模型更多地被作为基线模型和其他模型相比较。

我们一般不会在模型中假设所有协方差完全变动的情况，相反，我们会假设不同时间间隔的协方差为常数，或者假设不同时点间的相关是同一个值。这一较为严格的假设被称做“同质误差”，它等同于重复测量单变量方差分析（ANOVA）中的“复合对称性”假设。这一假设由于其简约化特征而十分具有吸引力，其模型总是使用最少的参数。然而，复合对称性这一强假设并不总是能够拟合现实世界的数据。

另一个常用于纵向数据的误差结构是自回归结构，它有时也被称做“一阶自回归”。该误差结构处于完全无限制与高度限制的同质误差假设之间，它假定误差项仅在一阶时间间隔之间相关。所以若间隔相关被估计为 0.30，则意味着时点 1 和时点 2 之间的相关、时点 2 和时点 3 之间的相关及其他类似项都为 0.30。在现实意义上，它表示时间间隔越大，则相关性越小。时点 1 和时点 3 之间的相关小于 0.30。自相关结构在简约性上几乎可媲美同质结构，它只需多估计一项参数——*rho*，即一阶相关。

表 3.8 中的拟合优度指标说明对于 LSOA 模型，自回归误差结构对数据的拟合最差，而无限制误差结构和同质误差结构则提供了大致相同的结果。但出于对模型简约性的考虑，我们将选择同质模型（事实上，表 3.7 的结果正是基于同质误差结构的）。之所以两个误差结构呈现了类似的结果，其原因在于我们的数据中仅有 4 个时点。当时点增加时，无限制模型的参数个数将显著上升，而不同误差结构的拟合结果也将显著不同。

本节只触及了在纵向数据分析中使用多层次模型的核心部分。若读者希望更多地了解诸如误差结构、非线性时变和非连续性时变以及其与时间序列分析的关系或是软件运用等问题，则可参考以下学者的著作：Little、Schnabel 和 Baumert(2000)；Moskowitz 和 Hershberger(2002)；Singer 和 Willett(2003)。

**表 3.8　不同误差结构下的模型拟合**

| 误差结构 | 参数 | 离差 | AIC | BIC |
|---|---|---|---|---|
| 无限制误差 | 15 | 73879 | 73909 | 74103 |
| 同质误差 | 12 | 73967 | 73991 | 74074 |
| 自回归误差 | 13 | 74340 | 74366 | 74456 |

# 附 录

## 数据及其他辅助材料

本书使用的两个数据集在以下网站上都可以下载：http://biostats. slu. edu/multimodel. htm。另外，该网站也提供用于分析这些数据并生成统计和图表结果的两个程序 HLM 及 R/S-Plus。读者还可在该网站获得多层次模型的其他资料。

## 软件

本书的重点在于提供多层次模型技术的理论及统计简介。尽管在本书的例子中，我们已经使用了两个常见的统计软件（HLM 和 R/S-Plus），但关于如何选择一个合适的软件以及某个软件的具体操作，却超出了本书的范围。并且任何的软件操作书籍在其出版后的短时间内即会过时。

对软件的选择或许是个艰难的任务，因为不同的软件会对多层次模型采取非常不同的拟合方法，并且各有其优势和劣势，而且某些软件可能非常昂贵。关于选择合适的软件，我们给出以下建议：

表 A1 对比了多层次模型常用软件的一些具体信息。多层次分析的初学者面临的一个问题可能是：应选择使用专业的多层次分析软件，还是选择运用一般统计软件中的混合效应程序（如 SAS 中的混合效应程序以及 R/S-Plus 中的 *lme* 或 *nlme*）？使用 SAS、SPSS 或 R 的主要优势是分析者很有可能已经拥有这类软件，并熟知其操作。然而一般来说，这些软件对多层次模型的具体分析而言，仍然缺乏足够的介绍。另外，这些软件在多层次模型的处理中都使用混

**表 A1　多层次模型软件相关信息**

| 多层次模型专门软件 | 版本 | 统计模型 | 层数 | GHLM | 界面 | 网　页 | 参考书目 |
|---|---|---|---|---|---|---|---|
| HLM | 5.04 | 多层次 | 3 | 有 | 菜单 | http://www.ssicentral.com/hlm/hlm.htm | Raudenbush 等人,2000 |
| MlwiN | 2.0(Beta) | 多层次 | 3+ | 有 | 菜单 | http://multilevel.ioe.ac.uk | Rasbash 等人,2000 |

| 一般统计软件 | 版本 | 统计模型 | 层数 | GHLM | 界面 | 网页 | 参考书目 |
|---|---|---|---|---|---|---|---|
| R/S-Plus-nlme or.lme | Lme4 0.4-4 | 混合 | 3 | 有 | 程序语句 | http://cran.r-project.org http://nlme.stat.wisc.edu/ | Pinheiro 与 Bates, 2000 |
| SAS-Proc Mixed | 8.2 | 混合 | 3 | 有 | 程序语句 | http://www.sas.com | Littell 等人,1996 Singer, 1998 |
| SPSS-MIXED | 12.0 | 混合 | 3 | 无 | 菜单或程序语句 | http://www.spss.com | SPSS Advanced Models documentation |

合效应模式，在模型较复杂的时候（如三层模型），这可能非常容易引起困惑。

HLM 和 MLwiN 都是处理复杂多层次模型的有效软件。它们的程序说明都比较清晰，尤其是 MLwiN，它提供了一套相当有用的程序讲解。在模型分析方面，它们相较一般软件限制更少，在模型收敛上不易出现问题，并且模型的拟合也相对更快速（尤其是 SAS，在估计复杂模型时非常慢）。但由于它们只是专业软件，因而在数据处理上有更多的限制，且绘图功能并不十分强大。

综合上述原因，若是进行大量且复杂的多层次模型分析，则至少我们应尝试使用 HLM 或 MLwiN。这两个软件在其主页上皆提供试用版。而上文的简短讨论仅关注了一些主要软件包，事实上，根据多层次模型的不同种类，我们还可使用大量其他的软件。下文列出了更多的相关信息。读者亦可参考 Leeuw 和 Kreft（2001）的著作，他们对多层次模型分析软件提供了更详细的比较。

## 其他资料

我们仍有其他大量的资料，它们或许可帮助初学者迅速掌握多层次分析技术，或许可帮助有经验的分析者处理更为复杂的多层次问题。多层次模型分析中心（the Centre for Multilevel Modeling）提供了一个可能是最好的网站：http://multilevel. ioe. ac. uk/。该网站除了提供对 MLwiN 的技术支持外，还提供完备的相关文献以及对其他多层次模型软件的介绍。

加州大学洛杉矶分校同样也有一个关于多层次模型的网站：http://statcomp. ats. ucla. edu/mlm/。它是一个非常有用的搜索引擎，能提供有关多层次模型的一切信息。其中特别值得关注的是 Singer 和 Willet 在《应用历时性数据分析》一书中使用的数据集和程序语句。读者可比较不同软件对同一多层次模型的处理方法，其中即包括我们上文提到的所有软件。

最后，浏览多层次模型讨论网对任何一位想要使用多层次分析技术的人都大有裨益。多层次模型的研究者和软件研发人员通常都会加入这一邮件列表。读者可以通过 http://www. jiscmail. ac. uk/lists/multilevel. html 加入列表并浏览过去的讨论。

# 参考文献

Akaike, H. (1987). "Factor analysis and the ATC." *Psychometrika 52*:317—332.

Armitage, P. & Colton, T. (1998). *Encyclopedia of biostatistics*. New York: J. Wiley.

Becker, R. A. & Cleveland, W. S. (1996). *S-Plus Trellis Graphics user's manual*. Seattle: MathSoft, Inc.

Bhaskar, R. (1989). *The possibility of naturalism: A philosophical critique of the contemporary human sciences*. Atlantic Highlands, New Jersey: Humanities Press.

Boyle, M. H. & Willms, J. D. (2001). "Multilevel modeling of hierarchical data in developmental studies." *Journal of Child Psychology and Psychiatry 42*:141—162.

Buka, S. L., Brennan, R. T., Rich-Edwards, J. W., Raudenbush, S. W. & Earls, F. (2003). "Neighhhorhood support and the birth weight of urban infants." *American Journal of Epidemiology 157*:1—8.

Carroll, K. (1975). "Experimental evidence of dietary factors and hormone-dependent cancers." *Cancer Research 35*:3374—3383.

Cleveland, W. S. (1993). *Visualizing data*. Summit, New Jersey: Hobart Press.

Curran, P. J., Stice, E. & Chassin, L. (1997). "The relation between adolescent and peer alcohol use: A longitudinal random coefficients model." *Journal of Consulting and Clinical Psychology 65*:130—140.

de Leeuw, J. & Kreft, I. G. G. (2001). "Software for multilevel analysis." In A. H. Leyland & H. Goldstein(eds.), *Multilevel modelling of health statistics* (pp. 187—204). Chichester, UK: Wiley.

Diez-Roux, A. Y, Merkin, S. S., Arnett, D., aet al. (2001). "Neighborhood of residence and incidence of coronary heart disease." *New England Journal of Medicine 345*:99—106.

Duncan, C., Jones, K. & Moon, G. (1998). "Context, composition and heterogeneity: Using multilevel models in health research." *Social Science and Medicine 46*:97—117.

Freedman, D. A. (2001). "Ecological inference and the ecological fallacy." In N. J. Smelser & P. B. Baltes(eds.), *International encyclopedia of the social & behavioral sciences*. New York: Elsevier.

Gebbie, K., Rosenstock, L. & Hernandez, L. M. (2003). "Who will keep the public healthy?" *Educating public health professionals for the 21st century*. Washington, D. C.: The National Academies Press.

Goldstein, H., Yang, M., Omar, R., Turner, R. & Thompson, S. (2000). "Meta-analysis using multilevel models with an application to the study of class size effects." *Applied Statistics 49*: 399—412.

Harrell, F. E. (2001). *Regression modeling strategies: With applications to linear models, logistic regression, and survival analysis*. New York: Springer.

Heck, R. H. & Thomas, S. L. (2000). *An introduction to multilevel modeling techniques*. Mahwah, NJ: Lawrence Erlbaum Associates.

Holmes, M. D., Hunter, D. J., Colditz, G. A., Stampfer, M. J., Hankinson, S. E., Speizer, F. E., Rosner, B. & Willett, W. C. (1999). "Association of dietary intake of fat and fatty acids with risk of breast cancer." *Journal of the American Medical Association 281*:914—920.

Hox, J. (2002). *Multilevel analysis*. Mahwah, NJ: Lawrence Erlbaum Associates.

Kreft, I. (1995). "The effects of centering in multilevel analysis: Is the public school the loser or the winner? A new analysis of an old question." *Multilevel Modelling Newsletter 7*:5—8.

Kreft, I. & de Leeuw, J. (1998). *Introducing multilevel modeling*. London: Sage Publications.

Lazarsfeld, P. F. & Menzel, H. (1969). "On the relation between individual and collecctive properties." In A. Etzioni(ed.), *A sociological reader on complex organizations*. New York: Holt, Rinehart, and Winston.

Leyland, A. H. & Goldstein, H. (2001). *Multilevel modeling of health statistics*. Chicester, UK: John Wiley & Sons.

Littell, R. C., Milliken, G. A., Stroup, W. W. & Wolfinger, R. S. (1996). *SAS system for mixed models*. Cary, NC: SAS Institute Inc.

Little, T. D., Schnabel, K. U. & Baumert, J. (2000). *Modeling longitudinal and multilevel data: Practical issues, applied approaches and specific examples*. Mahwah, NJ: Lawrence Erlbaum Associates.

Lochner, K., Pamuk, E., Makuc, D., Kennedy, B. P. & Kawachi, I. (2001). "State-level income inequality and individual mortality risk: A prospective, multilevel study." *American Journal of Public Health 91*:385—391.

Longford, N. T. (1989). "To center or not to center." *Multilevel Modelling Newsletter 1*:7, 11.

Longford, N. T. (1993). *Random coefficient models*. New York: Oxford University Press.

Luke, D. A. & Krauss, M. (2004, under review). "The influence of tobacco industry PAC contributions on voting behavior in the U. S. Congress."

Maes, L. & Lievens, J. (2003). "Can the school make a difference? A multilevel analysis of adolescent risk and health behaviour." *Social Science & Medicine 56*:517—529.

McArdle, J. J. & Epstein, D. (1987). "Latent growth curves within developmental structural equation models." *Child Development 58*:110—133.

Moos, R. H. (1996). "Understanding environments: The key to improving social processes and program outcomes." *American Journal of Community Psychology 24*:193—201.

Moskowitz, D. S. & Hershberger, S. L. (2002). *Modeling intraindividual variability with repeated measures data: Applications and techniques*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Mossholder, K. W., Bennett, N. & Martin, C. L. (1998). "A multilevel analysis of procedural justice context." *Journal of Organizational Behavior 19*:131—141.

Muthén, B. O. (1994). "Multilevel covariance structure analysis." *Sociological Methods & Research 22*:376—398.

O'Brien, R. M. (2000). "Levels of analysis." In E. G. Borgbatta & R. Montgomery(eds.), *Encyclopedia of sociology*(2nd ed.). New York: Macmillan.

"Office of Behavioral and Social Sciences Research."(2000). *Toward higher levels of analysis*:

*Progress and promise in research on social and cultural dimensions of healrh* (NIH Publication No. 01—5020). Bethesda, MD: National Institutes of Health.

Perkins, D. D., Wandersman, A., Rich, R. C. & Taylor, R. B. (1993). "The physical environment of street crime: Defensible space, territoriality and incivilities." *Journal of Environmental Psychology 13*:29—49.

Pinheiro, J. C. & Bates, D. M. (2000). *Mixed-effects models in Sand S-Plus*. New York: Springer.

Pinheiro, J. C., Bates, D. M., DebRoy, S. & Sarkar, D. (2003). *The nlme package*.

Plewis, I. (1989). "Comment on 'Centering predictors in multilevel analysis'." *Multilevel Modelling Newsletter 1*:6, 11.

Rasbash, J., Browne, W., Goldstein, H., Yang, M., Plewis, I., Healy, M., Woodhouse, G., Draper, D., Langford, I. & Lewis, T. (2000). *A user's guide to MLwiN* (Version 2. 1b). London: University of London, Institute of Education.

Raudenbush, S. W. (1989). "'Centering' predictors in multilevel analysis: Choices and consequences." *Multilevel Modelling Newsletter 1*:10—12.

Raudenbush, S. W. & Bryk, A. S. (1985). "Empirical Bayes meta-analysis." *Journal of Educational Statistics 10*:75—98.

Raudenbush, S. W. & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage Publications.

Raudenbush, S. W., Bryk, A. S., Cheong, Y. E. & Congdon, R. (2000). *HLM 5: Hierarchical linear and nonlinear modeling*. Lincolnwood, IL: SSI Scientific Software International.

Rice, N., Carr-Hill, R., Dixon, P. & Sutton, M. (1998). "The influence of households on drinking behaviour: A multilevel analysis." *Social Science & Medicine 46*:971—979.

Schwarz, G. (1978). "Estimating the dimension of a model." *Annals of Statistics 6*:461—464.

Shinn, M. & Rapkin, B. D. (2000). "Cross-level research without cross-ups in community psychology." In J. Rappaport & E. Seidman (eds.), *Handbook of community psychology*. New York: Kluwer Academic/Plenum Publishers.

Singer, J. D. (1998). "Using SAS PROC MIXED to fit multilevel models, hierarchical models, and individual growth models." *Journal of Educational and Behavioral Statistics 24*:323—355.

Singer, J. D. & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. New York: Oxford University Press.

Snijders, T. & Bosker, R. (1994). "Modeled variance in two-level models." *Sociological Methods & Research 22*:342—363.

Snijders, T. & Bosker, R. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. London: Sage Publications.

Villemez, W. J. & Bridges, W. P. (1988). "When bigger is better: Differences in the individuallevel effect of firm and establishment size." *American Sociological Review 53*:237—255.

# 潜变量增长曲线模型

## 作者简介

### 克里斯托弗·普里彻(Kristopher Preacher)

美国堪萨斯大学定量心理学系副教授。研究方向是运用因子分析、结构方程模型和多层模型来分析纵向与相关数据。另外,他还致力于开发检验调解与调节假设的方法、弥合理论与实践之间的差异以及研究多元方法应用到社会科学问题时的模型评估与模型选择。

### 阿斯荣·威克曼(Asron Wichman)

美国俄亥俄州立大学社会心理学课程博士后讲师。他的研究领域为社会认知与量化方法在个体差异方面,包括两者在性格评估上的应用。

### 罗伯特·麦卡勒姆(Robert MacCallum)

美国北卡罗来纳大学教堂山分校瑟斯通计量心理学实验室(L. L. Thurstone Psychometric Laboratory)主任,同时也是俄亥俄州立大学名誉教授。他的研究主要涉及分析相关数据的定量模型与方法,特别是因子分析、结构方程模型和多层模型。他还运用这些方法分析纵向数据,关注个体随时间变化模式的差异。

### 南希·布里格斯(Nancy Briggs)

澳大利亚阿德莱德大学公共卫生领域的统计学家,也是卫生与行为科学领域各种研究项目的数据分析师。她的研究是将高级多元统计方法(例如,线性或非线性多层模型和潜变量模型)应用于经验数据。

## 译者简介

### 姜念涛

曾获美国北卡罗来纳大学社会学和统计学硕士学位,目前是该校社会学系博士候选人。

何表达和估计的，并说明了它和多层次增长模型的关系。在本书中，作者显示了各种各样的模型，例如固定/随机截距和/或者固定/随机坡度、有条件增长曲线模型、平行过程模型、多组分析。他们也讨论了模型的各种延伸，例如分段增长、结构潜在增长、类型和有序变量结果等。他们用了大量说明性的图形来表述和讨论，这是使用结构方程模型的一个优势。

廖福挺

# 第1章 | 简介

社会和行为科学的一个根本观察就是人随时间而变化，但并不是以同样的方式或同样的速度。譬如，词汇能力在小学阶段会稳步地提高，但并不是所有的学生都会以同样的速度提高。在很多情况下，变化的速度和方向都存在个体差异，而这些变化中的个体差异往往具有科学上或实践上的意义。时间上的变化可以用秒（如在心脏反应研究中）或用十几年来测量（如在生命发展研究中）。例如，夫妻争执会在短短的几分钟内引起显著的内分泌变化，但这些变化对于丈夫和妻子而言，可能会十分不同（Kiecolt-Glaser et al.，1997）。Flora和Chassin(2005)检验了父母酗酒对青少年使用毒品增长的影响，跟踪调查项目参与者数年直到他们长大成人。早期研究变化的方法有很大的局限：(1)它们或只关注群体层面或只是个体层面的增长；(2)它们只着眼于两个时间点上的测量，导致数据过于贫乏而不能检验一些最基本的关于变化的有意思的假设。科学家常常需要更广泛的数据和更高级的统计方法来区分和理解变化的形态和方向（轨迹），并确定变化的原因和后果。

纵向设计能够得到关于心理现象的趋势和个体变化差异的有价值信息。这类数据的丰富性随着数据收集次数的增加而增加。Willett(1989)以及Willett与Sayer(1994)讨论了相对于二期数据，多期纵向数据的几个优势：(1)研究结果的质量会得到提高；(2)根据心理学理论可以建立正确的增长函数；(3)可以检验系统中个体间成长差异的假设；(4)可以将成长的特性与背景特点联系起来；(5)成长测度的精确性与准确性随着数据期数的增加而快速提高。简言之，纵向数据不仅提高了假设检验的统计效力，而且使研究者可以检验他们用横截面或二期数据无法检验的假设。

虽然利用纵向数据的这些优势使许多方法发展起来，但本书所关注的是潜变量增长曲线模型(LGM)。潜变量增长曲线模型代表了一类广泛的统计方法，这些方法允许更好的假设表述、提供更好的统计效力，并且比起别的方法，这些方法使统计模型与理论更好地衔接在一起。潜变量增长曲线模型不仅可以直接检测个人在不同时点的变化，而且能够检查不同个体之间的差异。这个模型的好处不仅在于它能够建构变化模型，还在于它允许研究者探寻变化的前因后果。

潜变量增长曲线模型框架可以让科学家清楚地表达和检验如下的问题：(1)随着时间的推移，均值变化趋势的形态是什么样的？(2)最初的水平是否可以预测变化速度？(3)两组或多组分组是否在变化轨迹上有区别？(4)是否可以通过均值变化趋势的速度或弯曲度预测重要的结果？(5)哪些变量与随着时间的推移而产生的变化有系统的关联？(6)根据观测数据，关于变化轨迹的理论假设是否站得住脚？(7)变化轨迹的形态是否存在显著的个体间差异？(8)一个变量的变化是否与另一个变量的变化相关？

上述问题当然不是详尽无遗的。接下来，我们将对本书的其余部分进行概述。

## 本书概述

本书是为那些对研究随时间推移的变化现象感兴趣的读者而著。不过，因为潜变量增长曲线模型是结构方程模型的一个应用，所以我们建议读者应具备基本的结构方程模型的知识。对于那些对潜变量增长曲线模型感兴趣的研究者而言，对多元线性回归、运用路径图来表示模型、模型识别问题以及固定参数和自由参数等话题有一定程度的熟悉是必要的。有用的入门书籍包括 Kline (2004)、Maruyama(1997)和 Raykov 以及 Marcoulides(2000)的著作。更高级的内容可以在 Bollen(1989)的书中找到。

为了演示潜变量增长曲线模型的功用和灵活性，我们首先从一个非常基本的模型开始，然后以此模型为基础进行常用的拓展。这些拓展允许在变化过程

中加入的更深入的理解和更清楚有力的假设检验。在每一个步骤中，我们都将模型应用于一个现成的数据集来演示研究者如何在应用中着手处理问题。我们会示范如何使用 LISREL(Jöreskon & Sörbom, 1996)、Mx(Neale, Boker, Xie & Maes, 2003)和 Mplus(L. K. Muthén & Muthén, 1998—2006)这 3 种流行的结构方程模型应用软件来拟合模型与数据，当然也可以使用其他方便使用的结构方程模型软件(譬如 EQS 和 AMOS)。由 LISREL 和 Mx 所获得的结果都是相同的(或几乎完全相同)。程序命令都列在我们的网站上[1]，但我们所描述的绝大部分的模型都可以通过几乎任何一个结构方程模型软件来实现。此外，我们还提供了一个详尽的参考文献列表，以方便有兴趣的读者获得更多的信息。所以，本书不仅使研究者能够了解潜变量增长曲线模型，而且可以帮助研究者深入探讨有关文献。

我们首先扼要地概述促使潜变量增长曲线模型发展的研究，然后描述正式的模型设定，随后是关于参数估计和模型评估的章节。在这些介绍性的材料之后，我们将介绍本书所使用的数据集和用于分析的软件。本书的其余部分描述了在实践中可能会遇到或用到的具体模型。从一个基本的(零)模型开始，我们通过一步步地放宽对参数的限制并添加另外的变量来探究更复杂的增长曲线模型。然后，我们探讨一些有趣的和常见的基本潜变量增长曲线模型的拓展，包括相关的增长曲线(多增长过程同时建模)、队列序列设计、添加随时间变化的协变量和更复杂的增长函数。在这些例子之后，我们将讨论潜变量增长曲线模型和一些其他技术的关系，包括增长混合模型、分段增长曲线、潜变量变化模型以及多层模型与潜变量增长曲线模型的区别。

## 潜变量增长曲线模型：简短的历史与概述

在历史上，增长曲线模型(例如 Potthoff & Roy, 1964)一直被用来为纵向数据建模。在纵向数据中，一些结果变量在不同时点被重复测量。潜变量增长曲线模型方法来源于探索性因子分析及相关文献。重复测量中的协变量可以用探索性因子分析(Baker, 1954; Rao, 1958)或主成分分析(Tucker, 1958、

1966)来建模。之后,因子或成分再被概念化为变化的方面或计时测量(与心理测量相对)因子(McArdle, 1989; McArdle & Epstein, 1987),因子负载可以被解释为重复测量对这些不可观测的变化方面依赖性的参数。这些变化方面可以包括线性、二次式或S形趋势。不过,这些方法用来研究变化时有许多问题。在实践中运用这些方法的一个主要障碍是因子旋转的不确定性——没有一个明确的因子旋转标准能选择一个负载模式来拟合可解释的变化方面(例如,一组多项式曲线)。虽然人们一直尝试发展可以用来识别平滑型函数的因子旋转标准(如 Arbuckle & Friendly, 1977; Tucker, 1966),但没有一个能完全令人满意。这些方法的另一个局限在于,它们是从估计未知函数趋势的自由负载的观点来为变化建模的(一种探索性方法),而不是检验一组特定的负载的可行性(一种验证性方法)。对实质性研究者而言,能够检验特定的假设趋势是非常重要的。

Meredith 和 Tisak(1990)描述了潜变量曲线分析——一个验证性因子分析的应用。这一分析通过允许研究者指定反映重复测量数据中特定的假设趋势负载,巧妙地回避了因子旋转不确定性的问题。这一潜变量曲线分析方法就相当于我们所说的潜变量增长曲线模型。因为潜变量增长曲线模型是验证性因子分析的一个应用,而后者又是结构方程模型的特例,所以增长曲线模型能够被放到更大的理论模型中。对潜变量增长曲线模型历史发展的细节感兴趣的读者,可以参考 Bollen 与 Curran(2006)为之提供的一个详尽的综述。

与其他竞争的方法(如协方差分析和多层模型)相比,潜变量增长曲线模型有几个优势。潜变量增长曲线模型允许对个体间在变化上的差异进行研究,并允许研究者研究变化的前因后果。潜变量增长曲线模型能够提供组级水平的统计指标,例如平均增长速度和平均截距,也能够对具体的曲线进行假设检验,并可以同时纳入随时间变化的变量和不随时间变化的变量。比起结构方程模型,潜变量增长曲线模型也有其优点,包括能够检验模型拟合指数、提供模型选择的标准、通过使用重复测量的潜变量解决测量误差问题,并能有效地处理缺失数据。不同总体和组别之间的比较也很简单明了。所以,潜变量增长曲线模型是一个非常灵活的模型策略,对于一些不太常见的情况也能够处理。

Curran 和 Willoughby(2003)在讨论增长曲线模型的时候阐述了一个观点,

认为增长曲线模型"可以被看成以变量为中心和以研究对象为中心的一个交叉点"(2003:603)。仅以变量为中心的变化视角强调平均趋势的变化,而仅以研究对象为中心的变化视角只聚焦于个人特征的变化趋势。当然每个视角都能够让我们获取重要的信息。与仅考虑变量或者研究对象为中心相比,潜变量增长曲线模型不仅能够获取平均的变化趋势,而且能够看出个人相较于平均的偏离。

## 模型设定和参数解释

潜变量增长曲线模型可以被看成结构方程模型的一个特例。结构方程模型是一个总的模型框架,它能够设定和检验不同变量之间的关系,其中有一部分变量是能观测到的,另一部分是潜变量。潜变量经常代表一些心理学概念,我们通常不能直接测量这些概念。一个典型的结构方程模型包含一小组潜变量,这些变量通常通过通径系数连接在一起,并被解释为回归权数。潜变量依次由可测量的标识变量来表示。潜变量和标识变量之间的关系就像因子分析模型。也就是说,因子负载代表了潜变量对标识变量的影响。

一般结构方程模型的特例可以生成基本的潜变量增长曲线模型。在这里,我们先展示这一特例,在后面,我们会应用更多一般化的结构方程模型来发展更复杂的潜变量增长曲线模型。在潜变量增长曲线模型中,变量 $y$ 是被重复测量的可测量变量,而潜变量的实质不是心理学的概念,相反,潜变量代表的是一种 $y$ 的变化模式或方面。在一个基本的潜变量增长曲线模型中,为了表示变化的方面,两个因子通常需要设定。这些因子通过设定重复测量变量 $y$ 的因子负载来定义,而这些因子负载描述了 $y$ 的变化趋势。截距因子描述当时间变量等于 0 时,结果变量 $y$ 的水平,而斜率因子则描述结果变量变化的速度。举例来说,一个对儿童外化的线性变化速度感兴趣的研究者可以收集外化行为的重复测量,然后把这些重复测量当做截距因子与斜率因子的标识(限制因子负载来反映对外化行为[①]变量变化模式的预期)。我们将会展示潜变量增长曲线模型

① 外化行为是心理学上的概念,一般指失控行为,如攻击、过失等行为问题。—— 译者注

框架的灵活性同样允许设定更为复杂的模型。

一个潜变量增长曲线模型可以依据数据模型、协方差结构和均值结构用矩阵表示法来表述数据模型描述因子(潜在变量)与 $y$ 的重复测量之间的关系。[2] 这个模型把 $p \times 1$ 观测值向量($\mathbf{y}$)表述为截距 ($\boldsymbol{\tau}_y$, $p \times 1$)、$m$ 个代表变化方面 ($\boldsymbol{\eta}$, $m \times 1$) 的潜变量以及扰动项($\boldsymbol{\varepsilon}$, $p \times 1$) 的一个线性函数,将因子负载($\boldsymbol{\Lambda}_y$, $p \times m$) 作为回归系数[3]:

$$\mathbf{y} = \boldsymbol{\tau}_y + \boldsymbol{\Lambda}_y \boldsymbol{\eta} + \boldsymbol{\varepsilon} \qquad [1.1]$$

$\boldsymbol{\tau}_y$ 项通常会因为模型识别的原因而被设定为 0。在展开式中 ($m = 2$),这一模型把 $y_{ti}$,即个体 $i$ 在时点 $t$ 的值,表述为两个潜变量($\eta_{1i}$ 与 $\eta_{2i}$)和一个扰动项($\varepsilon_{ti}$)的函数:

$$y_{ti} = \lambda_{1t}\eta_{1i} + \lambda_{2t}\eta_{2i} + \varepsilon_{ti} \qquad [1.2]$$

潜变量则可以表述为潜均值($\alpha_1$ 与 $\alpha_2$)和个体与这些均值偏差的函数:

$$\eta_{1t} = \alpha_1 + \zeta_{1t}\text{①} \qquad [1.3]$$

$$\eta_{2i} = \alpha_2 + \zeta_{2i} \qquad [1.4]$$

潜变量 $\eta_{1i}$ 与 $\eta_{2i}$ 常常被称为“随机系数”。表示个体与 $\eta_{1i}$ 与 $\eta_{2i}$ 的均值偏差的残差 $\zeta$ 往往被称为“随机效应”。

从方程 1.1 的数据模型中,我们可以导出一个协方差结构和一个均值结构。协方差结构将 $y$ 的重复测量的总体方差和协方差表示成模型参数的一个函数,而均值结构则将这些重复测量的总体均值表述为模型参数的另一个函数。均值结构和协方差结构与数据模型的区别在于,它们不包含个体观测对因子(如截距因子和斜率因子)的作用。这些模型常常被用来进行参数估计和模型评价。在协方差结构中,观测变量的方差和协方差矩阵 ($\boldsymbol{\Sigma}$, $p \times p$) 被表示为因子负载矩阵($\boldsymbol{\Lambda}_y$)、因子方差和协方差矩阵 ($\boldsymbol{\Psi}$, $m \times m$) 以及扰动项的方差和协方差矩阵 ($\boldsymbol{\Theta}_\varepsilon$, $p \times p$) 的函数(Bollen, 1989):

---

① 此处原文公式为 $\eta_{1t} = \alpha_1 + \zeta_{1t}$,应该是一处笔误。——译者注

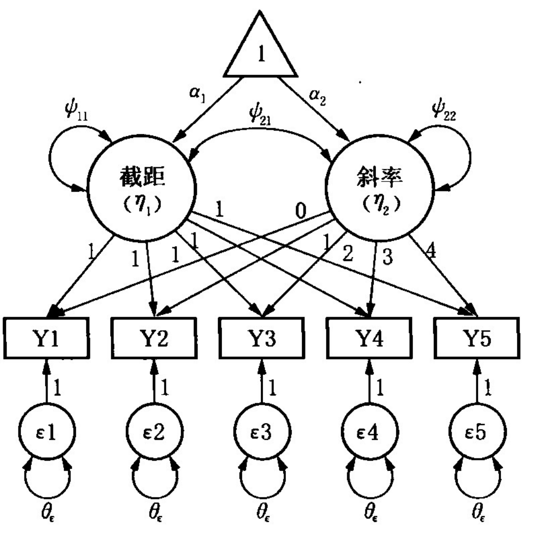
$$\mathbf{\Sigma} = \mathbf{\Lambda}_y \mathbf{\Psi} \mathbf{\Lambda}'_y + \mathbf{\Theta}_\varepsilon \qquad [1.5]$$

通过对数据模型求期望值而得到的均值结果，将观测变量 ($\boldsymbol{\mu}_y$，$p \times 1$) 的总体均值表示为截距 ($\boldsymbol{\tau}_y$，$p \times 1$) 和潜变量均值 ($\boldsymbol{\alpha}$，$m \times 1$) 的函数：

$$\boldsymbol{\mu}_y = \boldsymbol{\tau}_y + \mathbf{\Lambda}_y \boldsymbol{\alpha} \qquad [1.6]$$

在潜变量增长曲线模型中，$\boldsymbol{\tau}_y$ 的元素经常(但不是总是)被限制为 0，从而获得一个简化的数据模型和均值结构。因此，所考虑的参数就包含在矩阵 $\mathbf{\Lambda}_y$、$\mathbf{\Psi}$ 和 $\mathbf{\Theta}_\varepsilon$ 以及向量 $\boldsymbol{\alpha}$ 中。$\mathbf{\Lambda}_y$ 的列被称为“基础曲线”或“潜变量增长向量”(Singer & Willett，2003)。

在图 1.1 的模型中，Y1 到 Y5 表示变量 Y 的平均间隔的重复测量。在这里，Y 的变化被建模为两个基础曲线的函数，因此，负载矩阵 $\mathbf{\Lambda}_y$ 有两列。截距因子的负载被固定为 1 以表示一个常数对重复测量的影响，斜率因子的负载被固定为一个线性的级数以表示随时间变化的线性增长。虽然斜率负载一般被




注：根据结构方程模型的惯例，圆形表示潜变量，矩形表示测定变量(在这里，Y1 到 Y5 是平均间隔的重复测量)，三角形表示常数，双箭头表示方差或协方差，单箭头表示回归权数，数值对应固定参数，符号则表示自由参数(有待估计的参数)。

**图 1.1 一个典型的潜变量增长曲线模型的完整路径图**
**(包括随机截距、随机线性斜率和截距—斜率协方差参数)**

设定为从 0 开始，以使第一时点的测量表示初始反应，但这不是必须的，而且实际上往往是不恰当的。也可能存在额外的因子负载，其中每一个都表示增长的其他方面。这些负载常常（虽然并不一定）是多项式（参见第 2 章的模型 10）。此外，这些变化方面的协方差可以通过设定负载间的协方差路径来估计。估计这些协方差的能力在一些情况下是非常重要的，譬如当我们对确定一些变量的增长速度是否与其初始状态相关感兴趣时。图 1.1 中的三角形表示常数 1。因此，连接三角形与基本负载的路径系数就基于一个常数的回归，表示截距和斜率因子的均值。

在用矩阵表示 Y1 至 Y5 的协方差结构时，设定截距和斜率因子在重复测量上的负载等同于完全设定 $\mathbf{\Lambda}_y$，即因子负载矩阵的内容。对于一个具有同方差性和独立扰动协方差的简单线性增长模型，其中重要参数的必要限定包括：

$$\mathbf{\Lambda}_y = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{bmatrix} \qquad [1.7]$$

$$\mathbf{\Psi} = \begin{bmatrix} \psi_{11} & \\ \psi_{21} & \psi_{22} \end{bmatrix} \qquad [1.8]$$

$$\boldsymbol{\alpha} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} \qquad [1.9]$$

$$\mathbf{\Theta}_\varepsilon = \begin{bmatrix} \theta_\varepsilon & & & & \\ 0 & \theta_\varepsilon & & & \\ 0 & 0 & \theta_\varepsilon & & \\ 0 & 0 & 0 & \theta_\varepsilon & \\ 0 & 0 & 0 & 0 & \theta_\varepsilon \end{bmatrix} \qquad [1.10]$$

解释这些矩阵中的参数是直截了当的。研究者固定 $\mathbf{\Lambda}_y$ 中的元素来表示所

假设的变化轨迹，它的每一列负载表示假设的一个变化方面。在方程 1.7 中，所有第一列的元素都被设定为 1 来反映每个个体的截距在重复测量上保持不变这一事实。$\mathbf{\Lambda}_y$ 中第二列的线性级数表示相等时间间隔的线性增长假设。$\mathbf{\Psi}$ 的元素表示这些变化方面的方差和协方差。就图 1.1 中的简单线性增长来说，$\mathbf{\Psi}$ 矩阵包括截距方差($\psi_{11}$)、斜率方差($\psi_{22}$)以及截距与斜率的协方差($\psi_{21}$)。$\boldsymbol{\alpha}$ 的元素是从常数 1.0 来预测变化方面的回归系数，它们可以被解释为平均截距和平均斜率。特别是方程 1.9 中的 $\alpha_2$ 表示一个单位的时间变化所导致的结果变量的预期变化。最后，$\mathbf{\Theta}_\varepsilon$ 中的元素是扰动项的方差和协方差，表示数据中假设的潜变量曲线无法解释的方差部分。[4] 如果我们假定扰动项方差服从同方差性(如方程 1.10 所假定的)，单一的扰动项方差可以通过给 $\mathbf{\Theta}_\varepsilon$ 的对角线元素加上一个相等的限定条件来表示。$\mathbf{\Theta}_\varepsilon$ 中的非对角线元素常常被设定为 0 来表示不同时点的扰动项是不相关的，虽然这一假定并不是必要的。这个基本的潜变量增长曲线模型可以被拓展，比如加入截距和/或斜率的预测变量，多个结果变量的并行增长曲线以及参数限制(譬如，截距和斜率的协方差可以被限定为 0)等。相关拓展将在后面讨论。

概括来讲，不管重复测量的数目有多少，在一个典型的潜变量增长曲线模型的线性应用中，我们需要估计 6 个参数。这些参数包括平均截距和斜率($\alpha_1$ 和 $\alpha_2$)、截距和斜率的方差与协方差($\psi_{11}$、$\psi_{22}$ 和 $\psi_{21}$)以及一直保持不变的扰动项方差($\theta_\varepsilon$)。其余的参数通常被限定为 0 或与一个特定的变化模式相一致的数值。正如读者将会看到的，依赖于特定的应用所需，这些限定中的一部分可以通过各种方式被改变或放宽。在更为复杂的模型中，如预测非线性增长或变化的模型，其他自由参数还会被添加进来。

## 时间的测量尺度

在图 1.1 和方程 1.7 中，时间在某种程度上被重新编码以表示线性的变化，即将时间测量尺度的起点放在测量的第一时点 ($\lambda_{1,2} = 0$)。方程 1.7 中展示的负载模式，即截距被解释为“初始状态”，是典型的时间编码方式。但从特定的

研究问题出发，其他编码方式同样是有用的。如果在某种情况下，最后时点的测量（例如，从一个毒品康复计划中毕业）是研究最为关心的时点，把最后测量的时点编码为 0 就更为合理，如下面的负载矩阵所示：

$$\mathbf{\Lambda}_y = \begin{bmatrix} 1 & -4 \\ 1 & -3 \\ 1 & -2 \\ 1 & -1 \\ 1 & 0 \end{bmatrix} \qquad [1.11]$$

在方程 1.11 中，斜率的负载呈线性增长，但截距被定义为第五时点，即最后一个时点的均值。此外，如果用月而不是用年来编码时间则更为合理，方程 1.11 可以被重新参数化为：

$$\mathbf{\Lambda}_y = \begin{bmatrix} 1 & 0 \\ 1 & 12 \\ 1 & 24 \\ 1 & 36 \\ 1 & 48 \end{bmatrix} \qquad [1.12]$$

这一变化导致对方程 1.11 和方程 1.12 的负载矩阵中，与斜率因子相关的某些模型参数（如平均斜率或可观测标识变量对于斜率的回归权数）的解释有所不同，但整个模型的拟合程度不会改变。事实上，任何对负载矩阵的线性转化都不会改变模型拟合程度，虽然这些转化的确会对参数的解释产生重要的影响（Biesanz，Deeb-Sossa，Papadakis，Bollen & Curran，2004；Hancock & Lawrence，2006；Mehta & West，2000；Rogosa & Willett 1985；Stoel，2003；Stoel & van den Wittenboer，2003；Stoolmiller，1995；Willett，Singer & Martin，1998）。譬如，将时间变量对均值对中在很少的情况下是合理的，因为中央时点的情况常常不是我们所关心的。[5]更合适的做法是将时间变量编码为从某种程度上有助于回答实质问题的形式。时间测量尺度的零点应该被设在一个有意义的测量时点上，部分是因为这一选择决定了解释平均截距、截距方

差和截距—斜率协方差的时点(Rogosa & Willett, 1985)。Stoel(2003)以及Stoel与van den Wittenboer(2003)认为,截距只在存在自然原点的增长过程下才有意义。如果时间原点是任意选定的,那么应该避免对任何与截距相关的参数进行解释。同样,对于时间度量的选择应当便于增强可解释性(Biesanz et al., 2004)。在某些情况下,这可能会涉及在同一个模型中同时使用两个时间度量标准(例如,年龄和测量期,参见McArdel & Anderson, 1990)。

一个普遍的问题是关于所使用的重复测量的最优数目。这个选择在很大程度上是由实际的考虑,如时间与费用决定的。对$m$项多项式增长因子,无论不同时点的扰动项的方差是否被设定为相等,为了获得一个自由度至少为1的模型,最小的重复测量的数目应该是$m+1$。假如要估计的参数只包括因子均值、因子(协)方差和扰动项方差,这个公式就一直有效。一个自由度至少为1的模型(即一个过度识别的模型)是必须的,这是因为如果自由参数的个数等于或大于样本观测变量的均值和(协)方差的数目,模型就不能识别,也不能被检验(关于模型识别的更多细节,参见Bollen, 1989; Bollen & Curran, 2006)。

鉴于$m+1$是绝对的最小值,那么我们应当使用多少重复测量呢?Stoolmiller(1995)建议大概4至5个测量时点就足以为线性增长建模。MacCallum、Kim、Malarky和Kiecolt-Glaser(1997)认为并没有可靠的经验法则,但建议线性模型至少需要4至5个重复测量,而更为复杂的模型需要的测量可能会“比那多得多”(1997:217)。为了获得足够的统计效力来检验并行过程潜变量增长曲线模型(第2章的模型7)中的斜率因子间的协方差,Hertzog、Lindenberger、Ghisletta和von Oertzen(2006)推荐至少收集6个重复测量数据,但这个数目可以根据效应大小、样本数量或增长曲线的可靠性(总体方差被模型的增长方面所解释的比率)的变化而增加或减少。如同其他数据分析一样,更多就意味着更好,更多的数据产生更多的信息,这永远不会是一件坏事。不过,我们的经验是简约的线性模型在拟合超过6个重复测量时常常会有问题。这应该不是意外,因为很少有自然的变化过程会精确地遵循线性的模式。没有任何模型是正确的,因此研究者应当对随着信息的累积,简单的模型会拟合较差这一结果做

好心理准备。潜变量增长曲线模型最适合对大样本在有限个数的时点测量的趋势进行建模。无论测量的数目是多少，测量的范围应当足以跨越理论感兴趣的整个时间框架。

最后需要注意的是，测量的时点并不一定要平均隔开。譬如，纵向数据是在 1977 年、1979 年、1983 年和 1984 年收集的，测量时点之间的间隔就变成 2、4 和 1：

$$\boldsymbol{\Lambda}_y = \begin{bmatrix} 1 & 0 \\ 1 & 2 \\ 1 & 6 \\ 1 & 7 \end{bmatrix} \qquad [1.13]$$

方程 1.13 中的负载矩阵仍然表示线性的变化，斜率负载之间的间隔本身就是测量时点之间间隔的线性尺度转换。如果将 1977 年作为基准线，随后的测量时点就发生在 2 年、6 年和 7 年之后。在非线性模型（如第 2 章中描述的多项式潜变量曲线模型或第 3 章中讨论的结构潜变量曲线模型）中，收集 5 个以上的重复测量，并在变化发生最迅速的时期将测量设定得更加紧密常常是一个好主意。这有助于避免估计上的问题，并能对描述变化的参数进行更准确的估计。

我们的观点是，并没有一个通用的设计和相应的增长曲线模型。测量期的数目、测量期之间的时间间隔、时间的单位以及时间原点的设定在相同或不同的研究中都可能相当不同。更为深入的关于时间测量尺度的探讨可以在 Biesanz 等人（2004）、Curran 与 Willoughby（2003）、Hancock 与 Choi（2006）、Schaie（1986）和 Stoolmiller（1995）的著作中找到。

## 异步测量

$\boldsymbol{\Lambda}_y$ 中的负载与用户所熟悉的结构方程模型的其他固定参数有根本性的不同。无论时间是如何编码的，$\boldsymbol{\Lambda}_y$ 的内容都表示时间的函数。举例来说，斜率负

载包含在其他模型中可以被认为是时间自变量的数值中。许多潜变量增长曲线模型的应用都过于简单地假定对所有的个体而言，数据是在同一个时点收集的(Mehta & West, 2000)。这样的数据被称为“时间—结构数据”(Bock, 1979)。所有被研究的对象在相同测量时点的数值可以被放入一个共同的 $\mathbf{\Lambda}_y$ 矩阵，但是这在大多数应用中是不现实的。因为当个体并不是在同一时点被测量的，或该个体在同一时点但年龄不同的测量情况下，使用早先描述的基本潜变量增长曲线模型就不恰当了。如我们将要讨论的，当个体在不同时点被测量时，主要有两种策略来估计增长曲线模型。这两种策略都允许个体具有不同的负载因子。

当并不是所有的个体都在同一时点被测量，但是有一套有限的测量时间表时，我们就可以运用多群体策略，即同时估计多个有不同的 $\mathbf{\Lambda}_y$ 矩阵的模型。我们将在下一节更详细地讨论这一策略，但基本上它涉及将有共同测量时间表的个体分组，然后对所有的组同时进行模型拟合(Allison, 1987; T. E. Duncan, Dunca, Strycker, Li & Alpert, 1999; McArdle & Bell, 2000; McArdle & Hamagami, 1991; B. Muthén, Kaplan & Hollis, 1987)。因此，所有只在时点1、时点3、时点5和时点6被测量的个体就属于第一组，而所有只在时点1、时点3、时点4和时点5被测量的个体就属于第二组。

如果在有太多不同的测量时间表的情况下，多群体的办法就不能解决问题了，但有一个更普遍的解决办法。Mx 和 Mplus 程序的一个引人注目的优点就是它们可以通过运用定义变量或个体数据矢量——包含每个个体固定值的特殊矢量——来包含所有个体的斜率负载(Hamagami, 1997; Neale et al., 2003)。传统的方法是将一个模型应用到被假定有同样的基本曲线和同样的 $\mathbf{\Lambda}_y$ 的所有个体上，定义变量的运用则是给每个个体都建立一套独特的斜率因子负载。在涉及年龄的纵向研究的特殊情况中，这被称为“对个体调整年龄”(Mehta & West, 2000)。Mehta 与 West(2000)的著作中的附录 A 和我们的网站(http://www.quantpsy.org)上都包括了展示这一方法的 Mx 程序样例。

## 前提假设

使用最大似然估计的潜变量增长曲线模型需要一些重要的前提假设。大多数前提假设涉及潜变量的分布(例如潜变量增长曲线模型中的截距、斜率和扰动项等)。因为根据定义,这些变量是不可观测的,根据惯例就需要假定它们的特征。我们假定方程1.2、方程1.3和方程1.4中,扰动项与残差的均值为0。在每个时点,这一假设不仅针对总体均值,而且针对同一个体在理论情况下重复观测的均值。也就是说,如果能够在一个特定时点对同一个个体重复测量,我们就假定这些量度的扰动项的均值为0。同样,所有残差项的同一时点或时点之间的协方差都被假定为0,方程1.2中,残差与随机截距和随机斜率的协方差也被设定为0。为了使用最大似然估计法,我们还需要假定可观测的变量是从一个与多元正态分布有大致相同的峰度的总体分布中得来的。潜变量增长曲线模型的前提假设在Bollen与Curran(2006)的著作中有大量的论述。

Byrne与Crombie(2003)讨论了另外3个前提假设。他们要求假设轨迹是线性的、不同时点的扰动项互不相关以及所有时点的扰动项的方差不变。实际上,这些并不是潜变量增长曲线模型或最大似然估计的前提假设。Byrne和Crombie的线性假设指的是增长轨迹的线性,而不是负载。这其实不是一个前提假设,而是要仔细检查的中心假设。两种可以检验线性假设的方法是比较线性增长曲线模型的拟合度与一个绝对拟合的基准线模型或一个未指定轨迹的模型的拟合度(见第2章的模型11)。Byrne和Crombie讨论的另外两个前提假设(独立性与扰动项的同方差性)可以是潜变量增长曲线模型设定的通常方面,但并不是必须的。事实上,能够估计不同特定时点的扰动项方差被认为是潜变量增长曲线模型方法的一个长处,也是将潜变量增长曲线模型与自回归策略相结合的方法所要求的(Curran & Bollen, 2001; McArdle, 2001)。

## 参数估计与缺失数据

传统的结构方程模型的参数估计是通过最大似然估计来实现的，使用这一方法时，需要假定测量的变量符合多元正态分布。对诸如潜变量增长曲线模型这类解释测量变量的协方差和均值的模型而言，数据常常是以从全部数据（数据有时不得不通过删除某些观测值或对某些缺失值进行估算的方法来保持完整）计算得来的样本协方差矩阵 $\mathbf{S}$ 和样本均值向量 $\bar{\mathbf{y}}$ 的形式表示。矩阵 $\mathbf{S}$ 有 $p\times p$ 阶，包含 $y$ 的 $p$ 个重复测量的样本方差与协方差。向量 $\bar{\mathbf{y}}$ 包含 $p$ 个重复测量的样本均值。根据方程 1.5 和方程 1.6 的协方差与均值结构，总体协方差矩阵 $\boldsymbol{\Sigma}$ 和均值向量 $\boldsymbol{\mu}$ 是模型参数的函数。如果我们把方程 1.5 和方程 1.6 中所有的参数放到一个向量 $\theta$ 中，那么参数估计的目标就是找到 $\hat{\theta}$ 中的参数估计值，从而使引申矩阵 $\boldsymbol{\Sigma}$ 和 $\boldsymbol{\mu}$ 同 $\mathbf{S}$ 和 $\bar{\mathbf{y}}$ 分别最为相近。在最大似然估计法中，这一最优性是用多元正态概率函数来定义的，即由最大似然估计法产生的一组参数估计使似然函数的对数最大化时的 $\hat{\theta}$ 为：

$$\ln L = -\frac{1}{2}\sum_{i=1}^{N}\{p\ln 2\pi + \ln|\boldsymbol{\Sigma}| + (\mathbf{y}_i - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{y}_i - \boldsymbol{\mu})\} \qquad [1.14]$$

Jöreskog(1967)将这一发展推广，可以证明 $\ln L$ 在下列差异函数最小化时得以最大化：

$$F_{ML} = \ln|\boldsymbol{\Sigma}| - \ln|\mathbf{S}| + \mathrm{tr}[(\mathbf{S}-\boldsymbol{\Sigma})\boldsymbol{\Sigma}^{-1}] + (\bar{\mathbf{y}}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\bar{\mathbf{y}}-\boldsymbol{\mu}) \quad [1.15]$$

因此，根据给定的 $\mathbf{S}$ 和 $\bar{\mathbf{y}}$，最大似然估计法寻找一个参数估计值的向量 $\hat{\theta}$，由这一向量产生的暗含矩阵 $\boldsymbol{\Sigma}$ 和 $\boldsymbol{\mu}$（见方程 1.5 和方程 1.6）使 $F_{ML}$ 最小化。请注意，如果 $\boldsymbol{\Sigma} = \mathbf{S}$ 而 $\boldsymbol{\mu} = \bar{\mathbf{y}}$（即模型完美地复制了数据），那么 $F_{ML}$ 等于 0。

差异函数 $F_{ML}$ 的最小化估计方法假定所有的样本数据都用来获得 $\mathbf{S}$ 和 $\bar{\mathbf{y}}$。如果有的数据缺失了（例如，个体是在不同时点测量的或某些个体在某些时点的数据未能得到），那么这一策略就不起作用了，因为从可用的数据计算得出的协方差矩阵不一定是内部一致的。幸好，我们有处理缺失数据的方法。上文提

到的在多个测量时间组合情况下，多群体策略可以被看做处理缺失数据的一个一般的、基于模型的方法（T. E. Duncan & Duncan, 1995; Marini, Olsen & Rubin, 1979; McArdle & Hamagami, 1992; B. Muthén et al., 1987）。举例来说，如果某些个体是在时点 1、时点 2、时点 3 和时点 5 测量的而其他个体是在时点 1、时点 3、时点 4 和时点 5 测量的，那么可以设定一个两组模型，其中的每一组成员都有共同的测量时点。对于多样本分析，方程 1.15 中的差异函数可以被推广为一个多样本的表达式并被最小化，以使所有组同时获得最优拟合。

在存在许多不同测量时间表的情况下，处理缺失数据的多组方法变得不现实。直到最近，这一局限一直是一个实实在在的问题，因为实际中的纵向设计往往含有缺失数据。估计方法的发展允许模型被直接用于拟合原始数据，这使得在分析中包括不完整的个案变为可能。在涉及缺失（不完整）数据的情况下，完全信息最大似然（FIML）估计方法常常被推荐用于获得最大似然参数估计值。为适用于不完整数据，方程 1.14 中表示完整数据的似然对数可以被修改为：

$$\ln L = -\frac{1}{2}\sum_{i=1}^{N}\{p_i \ln 2\pi + \ln|\boldsymbol{\Sigma}_i| + (\mathbf{y}_i - \boldsymbol{\mu}_i)'\boldsymbol{\Sigma}_i^{-1}(\mathbf{y}_i - \boldsymbol{\mu}_i)\} \qquad [1.16]$$

$\mathbf{y}_i$ 是个体 $i$ 的数据向量中可测量的部分，$\boldsymbol{\mu}_i$ 和 $\boldsymbol{\Sigma}_i$ 分别是模型的均值向量和协方差矩阵，其行和列对应个体 $i$ 的数据（Arbuckle, 1996; Wothke, 2000）。FIML 估计法需要最大化这一函数。FIML 运用的数据估算或删除部分数据的方法更为有效而且偏差较少，并能在数据完全随机缺失（MCAR）或随机缺失（MAR）的情况下获得无偏估计值（Neale, 2000; Rubin, 1976）。[6] 与配对删除、列删除和缺失值估算比较，FIML 已经被证明具有更低的收敛失败率、更高的效率、较少的偏差和更准确的模型拒绝能力（Enders & Bandalos, 2001）。这两种方法的特性相似。实际上，当没有数据缺失时，它们是相等的。更清楚与深入的、关于 FIML 算法及其与其他模型相比较的描述，参见 Enders(2001)的著作。

FIML 只是在出现缺失数据的情况下可以通过最小化差异函数来获得参数估计值的几种方法之一。其他方法包括将广义最小二乘法、未加权最小二乘法、E-M 算法和渐进自由分布法运用于通过配对删除、列删除、单一或多重缺失

值估算而转化来的“完整”数据。与这些方法相比，FIML 常常是更好的。这是因为：(1)它用所有可获得的信息来估计参数；(2)它不需要非常大的数据；(3)标准误可以由参数估计值的渐进协方差矩阵的逆矩阵获得。配对删除与列删除略去了一些数据，而配对删除更增加了碰到非正定的协方差矩阵的风险。此外，FIML 现在已经是大多数 SEM 软件，包括 AMOS(Arbuckle & Wothke, 1999)、Mplus(L. K. Muthén & Muthén, 1998—2006)、Mx(Neale et al., 2003)、EQS(Bentler, 1995)和 LISREL(Jöreskon & Sörbom, 1996)的标准估计选项。例如，在 LISREL 中，如果原始数据是输入数据，则 FIML 会被自动调用。缺失数据的问题在 Allison(1987、2002)中有更深入的描述。使用 FIML 估计法的主要缺点在于，如果有些数据缺失，就得不到所有 ML 的拟合指标。[7]

## 模型评估与选择

设定和检验模型作为理论预测是现代经验科学实践的基础。在发展一个模型以拟合观测的数据时，关键在于设定的模型要准确地反映增长的具体理论预测或意涵(Collins, 2006; Curran, 2000; Curran & Hussong, 2003; Curran & Willoughby, 2003)。此外，虽然这里的关注点是模型评估，但更好的做法往往是比较有竞争力的、源自理论的模型，而不是孤立地设定和评估模型。给定一个理论上可行的模型，LGM 可以通过评估包括截距和斜率因子的均值以及变化方面的方差与协方差等模型参数的统计或实践显著性来检验该模型。一个关于参数显著性的非正式的检验是用点估计值除以它的标准误，如果这一比率超过 2.00(1.96 在大样本的情况下)，那么这个参数估计值就被称为“在 0.05 的置信水平上与 0 显著不同”。

在 SEM 中，整体模型的拟合度也可以被评估。实际上，好的总体拟合度常常是理解参数估计值的先决条件。[8]在多元正态分布和完全拟合的原假设下，$\hat{F}_{ML} \times (N-1)$ 是一个自由度为 $[p(p+3)/2]-q^*$ 的卡方分布，$p$ 是变量的数目，而 $q^*$ 是自由模型参数的有效个数。[9]这个卡方值是一系列用来估计模型的预测值与观察数据之间匹配度的拟合指标的基础。除此之外，我们推荐在几个

SEM 程序中都有的近似误差均方根法(RMSEA)(Browne & Cudeck, 1993; Steiger & Lind, 1980):

$$\text{RMSEA}=\sqrt{\frac{max\left\{\left(\hat{F}_{ML}-\frac{df}{N-1}\right),\ 0\right\}}{df}} \qquad [1.17]$$

方程 1.17 中根式下的分子是总体模型不适度(拟合差异)的估计。因此,根式下的数量表示估计的每个自由度的总体模型误差,所以值越小就越好。RMSEA 是总体不适度的估计而不是简单的样本不适度的测量,因此被认为是较好的指标。重要的是,我们可以获得 RMSEA 的置信区间,为这一拟合指数提供点估计之外的一个精确度的测量。

因为以卡方为基础的指数常常存在于近乎不可逆矩阵(Browne, MacCallum, Kim, Andersen & Glaser, 2002)、违反分布的假设(Curran, West & Finch, 1996)和大样本数(Tucker & Lewis, 1973)相关的问题,所以最好检查 **S** 与 $\hat{\mathbf{\Sigma}}$ 对应元素间的简单残差。一个常用的、只基于残差的指数是标准化残差均方根(SRMR)(Jöreskog & Sörbom, 1996),即一个关于残差大小的概括量度。SRMR 是观测相关与模型方差和协方差矩阵所显示的相关之间的绝对差的平方均值的平方根(因此是越小越好)。同 RMSEA 一样,SRMR 也被许多 SEM 软件列为默认的输出,但要记住,SRMR 只评估协方差结构的拟合度。

另一类拟合指标反映的是一个特定的模型与一个恰当设定的空模型(见第 2 章的模型 0)之间的拟合度差异。这类拟合指标的一个例子是非基准拟合指数(NNFI)(Bentler & Bonett, 1980; Tucker & Lewis, 1973):

$$\text{NNFI}=\min\left\{\left(\frac{\frac{\chi_0^2}{df_0}-\frac{\chi_k^2}{df_k}}{\frac{\chi_0^2}{df_0}-1}\right),\ 1\right\} \qquad [1.18]$$

$\chi_0^2$ 与 $df_0$ 是由空模型计算而来的,而 $\chi_k^2$ 与 $df_k$ 则是在研究模型的基础上计算而来的。NNFI 已经被证明在分布假设被违反的情况下仍然比较稳健(Lei & Lomax, 2005)。我们在下一章将详细讨论恰当的空模型。在本书中,我们为所

有拟合的模型报告卡方、RMSEA、NNFI 和 SRMR 估计值。

除了孤立地评估模型，也可以用一种模型选择策略来评估嵌套模型或非嵌套模型的相对拟合度。如果一个模型(模型 A)的估计参数是另一个(模型 B)的估计参数的子集，则称前者嵌套在后者中。也就是说，如果对 B 中的某些参数加以限制而得到 A，那么 A 就嵌套在 B 中。当一部分数据缺失时，有的拟合指标，如 GFI 与 SRMR，就不能计算了(Enders, 2001)。在模型之间没有区别的原假设下，完全数据时的卡方值差异($\Delta\chi^2$)或不完全数据时 $-2\ln L$ 值的差异本身符合卡方分布，其自由度等于要估计的参数数目的差异。对于非嵌套的模型，利用信息的模型选择方法(如赤池信息准则、贝叶斯信息准则)可以用来选择模型。

## 统计效力

同绝大多数推论统计的应用一样，统计效力在 LGM 中是很重要的。效力指的是正确否决一个错误的原假设的概率。在 LGM 中，原假设是研究者的潜变量增长曲线模型，所以效力就是当研究者的增长模型不适用于总体情况时，它被否决的概率。LGM 模型往往在总体中不都是那么正确，所以一个高水平的统计效力会导致对非常好但并不完美的模型的否决。这种对好的模型的否决在实践中当然是不受欢迎的，但这是似然比检验模型拟合度的一个众所周知的局限。在实践中，为了补救这种情况，可以使用前面所讨论的各种描述性拟合测度，而不是仅仅专注于似然比检验模型拟合度。

MacCallum、Browne 与 Sugawara(1996)描述了一种涉及 RMSEA 拟合指标的统计效力(给定样本量)或最小必须样本量(给定所需的效力水平)的估算方法。研究者选择分别反映好拟合($\varepsilon_0$)与差拟合($\varepsilon_A$)的 RMSEA 值的原假设和备择假设。例如，完全拟合的检验可能会选择 $\varepsilon_0 = 0.00$ (完全拟合)和 $\varepsilon_A = 0.08$ (一般拟合)。而一个严格拟合的检验可能涉及 $\varepsilon_0 = 0.05$ (精密拟合)和 $\varepsilon_A = 0.10$ (不可接受拟合)。作者提供的 SAS 程序会提供在一个给定的效力水平上拒绝一个差模型的最小必须样本量 $N$、模型自由度 $df$ 以及用 RMSEA 表

示的代表好拟合与差拟合的原假设及备择假设。与之相反，他们的程序也提供在给定的 $N$ 和模型自由度的情况下，一对原假设与备择假设相对应的统计效力水平。

除了拒绝一个差模型的效力之外，我们也应当考虑检验非零参数的效力。这类效力是 LGM 中一个几乎没有研究过的课题。一个例外是 Hertzon 等人(2006)最近的一个研究。他们发现检验并行过程潜变量增长曲线模型(见第 2 章的模型 7)的斜率协方差的功效在很大程度上依赖于效应大小、重复测量的次数、增长曲线的可靠性以及样本量。

# 第2章 将潜变量增长曲线模型应用于经验数据

## 数据

接下来，我们将示范在实践中如何使用增长曲线模型。在示范中，我们将使用国家儿童健康与人类发展研究院(NICHD)的早期儿童照料与青年时期发展研究(NICHD早期儿童照料研究网络，2006)的数据。将研究问题量化之后对父母—子女关系的分析为解决类似的问题提供了一个相对客观的方法。大量的研究检验了这些亲子关系与诸如吸烟或饮酒(如 Blackson, Tarter, Loeber, Ammerman & Windle, 1996)、自我约束能力(如 Wills et al., 2001)和青少年怀孕(Miller, Benson & Galbraith, 2001)等结果的联系。研究发现良好的父母—子女关系与儿童正面的发展相关联，但某一时刻的父母—子女关系与其后的测量结果的相关性往往不是很强。父母—子女关系缺乏理想的有效预测，其诸多原因中的一个就是这些关系随时间而变化。心理动力学(Freud, 1958)、进化理论(Steinberg, 1989)和社会—认知(Smetana, 1988)的各种理论观点都预测随着子女的成熟，父母—子女关系将会发生变化，尽管变化的原因不同。研究父母—子女关系随时间的变化为增强预测准确度提供了额外的机会，并且能增强我们对父母—子女的互动如何影响儿童和青少年的不良行为的理解。

我们从父母—子女关系量表中挑选了两个复合测度。具体来说，15个学生—教师关系量表(Pianta, 1993)中的等级量表项目被用来评估子女对父母依恋的程度。每一个项目都在1到5的量表上打分，1表示“完全不适用”，而5表

示"完全适用"。这些项目被用来建构 4 个复合测度,分别是父母与子女的冲突度(CNFL)和亲密度(CLSN)。这些变量是在小学一年级到六年级(二年级时没有收集数据)时测量的。母亲—子女亲密度是绝大多数将要展示的模型中的主要重复测度变量。与这些关系测度一起,一个组群变量(子女的性别)被选中,以便我们展示如何在纵向模型中包括组群。我们抽出所有至少有一个有效 CLSN 分数的儿童的数据,得到一个包含 1127 个儿童(571 个男孩与 556 个女孩)的样本并分析所有在所关心的变量上有完整数据的个案。[10]表 2.1 给出了描述性的信息;图 2.1 给出了 CLSN 的平均分数;最后,表 2.2 给出了样本中 851 个完全数据个案的观测协方差矩阵与均值。

**表 2.1 母亲对子女亲密度数据的描述统计**

| 年级 | 整个样本 | | | 男孩 | | | 女孩 | | |
|---|---|---|---|---|---|---|---|---|---|
| | 样本数 | 均值 | 标准差 | 样本数 | 均值 | 标准差 | 样本数 | 均值 | 标准差 |
| 不完整数据的案例 | | | | | | | | | |
| 1 | 1016 | 37.96 | 2.56 | 508 | 37.73 | 2.77 | 508 | 38.20 | 2.30 |
| 3 | 1025 | 37.19 | 2.82 | 508 | 37.07 | 2.73 | 517 | 37.32 | 2.90 |
| 4 | 1022 | 36.96 | 3.33 | 516 | 36.62 | 3.53 | 506 | 37.31 | 3.08 |
| 5 | 1018 | 36.56 | 3.25 | 506 | 36.65 | 3.34 | 512 | 36.77 | 3.16 |
| 6 | 1024 | 36.18 | 3.56 | 512 | 35.93 | 3.63 | 512 | 36.42 | 3.47 |
| 完整数据的案例 | | | | | | | | | |
| 1 | 851 | 37.95 | 2.53 | 417 | 37.76 | 2.66 | 434 | 38.14 | 2.38 |
| 3 | 851 | 37.28 | 2.74 | 417 | 37.20 | 2.62 | 434 | 37.35 | 2.86 |
| 4 | 851 | 37.05 | 3.28 | 417 | 36.74 | 3.45 | 434 | 37.34 | 3.07 |
| 5 | 851 | 36.57 | 3.21 | 417 | 36.34 | 3.31 | 434 | 36.79 | 3.09 |
| 6 | 851 | 36.14 | 3.59 | 417 | 35.84 | 3.73 | 434 | 36.42 | 3.43 |

**表 2.2 母亲—子女亲密度:完整数据的均值与协方差 ($N$ = 851)**

| 年级 | 均值 | 协方差 | | | | |
|---|---|---|---|---|---|---|
| 1 | 37.9542 | 6.3944 | | | | |
| 3 | 37.2785 | 3.2716 | 7.5282 | | | |
| 4 | 37.0463 | 4.1435 | 6.0804 | 10.7290 | | |
| 5 | 36.5696 | 3.7058 | 5.1597 | 6.5672 | 10.2920 | |
| 6 | 36.1363 | 4.1286 | 5.7608 | 7.2365 | 7.6463 | 12.9085 |

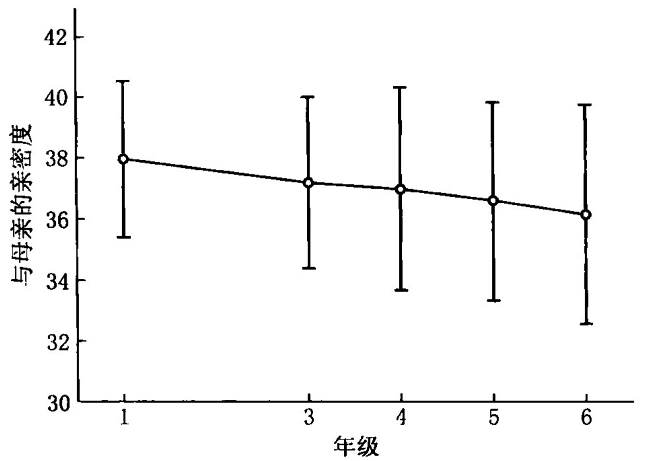



**图 2.1　平均亲密度分数(误差条代表一个标准差的宽度)**

## 软件

以下展示的模型是用 3 种软件包来估计的。我们使用 LISREL 8.8 (Jöreskog & Sorbom, 1996)、Mx 1.1(Neale et al., 2003)和 Mplus 4.2(L. K. Muthén & Muthén, 1998—2006)。还有其他易于分析增长曲线模型的软件包，如 AMOS 与 EQS。其中，作为一个灵活、没有版权限制并能免费下载的软件，Mx 特别实用。在本书写作时，能够估计本书中所有模型的一个免费学生版的 LISREL 可以从 SSI 的网站上获得。[11] Ferrer、Hamagami 与 McArdle(2004)提供了一个在各种软件中设定增长曲线模型的指南。

## 模型拟合策略的概要

一个对重复测度问题的 LGM 的典型应用包含变量在两个分析层上的测量。第二层单位是研究的实体常常是(但并不一定是)个人。第一层单位是在每一个第二层单位上获取的重复测度。其他变量可以是在第一层或第二层上的测度。第一层变量包括结果变量以及其他在同一时点测量的变量。被重复测量并用于预测所有结果变量、同时在重复测度之间具有差异性的变量叫

做“时变协变量”(TVCs)。第二层的变量表示第二层单位的特性,因此它们在个人间不同而不随时间变化。这样的例子包括性别、环境变量或稳定的性格特征。第二层的自变量常常被称为“非时变协变量”。截距或斜率因子的变化性可能会由非时变协变量来解释。举例来说,一个研究者可能对随着儿童的成长,父母对母亲—子女亲密度的理解的变化感兴趣。研究者可能在某几个时点从同一批父母那里收集数据并尝试为随时间变化的亲密度建模。对于这一模型,这位研究者可以选择将时间标度的零点放在第一个测量时点,然后再对同一批儿童收集 4 次数据。如果研究者发现不同儿童的轨迹(截距与斜率)明显不同,可能会试图研究是否性别差异能够解释部分的个体间差异性。因此,这位研究者可以引入性别作为最初水平(截距因子)和变化速度(斜率因子),或两者共同的一个自变量。在这里,性别是一个非时变协变量,因为它在第二层单位(儿童)之间而不在第二层单位内部变化。

特别对于像 LGM 这样灵活的方法,在尝试用模型拟合数据之前,设定一系列根植于理论的、用于检验的模型是有帮助的。然后我们可以按照特定的顺序,检查一系列先验的、理论上合理的模型。当更复杂的模型已经不再能显著地提高模型的拟合度时,我们就得出以下结论:已经找到一个可接受的模型。在下一节中,我们将展示这一过程。

## 模型 0:空模型

“空模型”这一说法通常是用来指比较不同假设模型的一个基础。LGM 中的空模型与典型的 SEM 应用中的空模型不同。在典型的 SEM 应用中,空模型是一个测量变量之间没有任何预测关系的模型,模型中也没有潜变量,只有方差参数被估计。然而,在 LGM 的背景下,我们把空模型定义为一个随时间推移而没有变化并在总体平均水平上没有差异性的模型(Widaman & Thompson, 2003)。只有平均水平(截距 $\alpha_1$)和一个共同的扰动项方差($\theta_\varepsilon$)被估计:

$$\mathbf{\Lambda}_y = \begin{bmatrix} 1 \\ \vdots \\ \vdots \\ 1 \end{bmatrix} \qquad [2.1]$$

$$\mathbf{\Psi} = [0] \qquad [2.2]$$

$$\boldsymbol{\alpha} = [\alpha_1] \qquad [2.3]$$

$$\mathbf{\Theta}_\varepsilon = \begin{bmatrix} \theta_\varepsilon & & & \\ 0 & \ddots & & \\ & & \ddots & \\ 0 & & 0 & \theta_\varepsilon \end{bmatrix} \qquad [2.4]$$

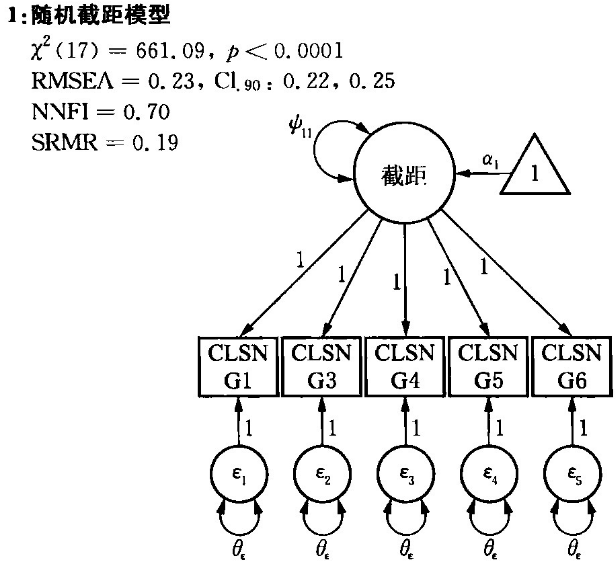
这里，$\mathbf{\Lambda}_y$ 是一个表示每一个我们测量时点截距因子的固定负载的 $5 \times 1$ 矩阵。假定这个模型随时间推移不发生变化，所以斜率因子就被省略了。矩阵 $\mathbf{\Psi}$ 是一个包含截距方差的 $1 \times 1$ 矩阵，在这一模型中，截距方差被设定为 0。$\mathbf{\Theta}_\varepsilon$ 是一个所有对角元素被限定为相等的 $5 \times 5$ 对角矩阵。这一相等的限定表示同方差性的假设。[12]最后，$\boldsymbol{\alpha}$ 是一个包含估计的整体均值 $\alpha_1$ 的 $1 \times 1$ 矩阵。如果空模型不适合数据（它常常是这样的），常常要估计截距的方差，并将一个线性的斜率因子纳入模型以表示随时间推移发生的变化。根据一个先验的理论指导或一个更具探索性的方法，截距和斜率的自变量也可以被纳入。这个简单的两参数的模型或一个与它尽可能相近的模型将作为本章其他部分计算 NNFI 时的空模型。

## 模型 1：随机截距模型

随机截距模型是潜变量增长曲线模型的一个最简单的例子。在 LGM 中，随机截距模型相当于一个加入均值结构的单因素验证性因子分析（CFA）模型，其中的所有因子负载都被设定为 1，而所有的扰动项方差都被限定为相等（见图 2.2）。参数的矩阵同方程 2.1、方程 2.2、方程 2.3 与方程 2.4 中设定的一

样，只是 $\Psi = [\psi_{11}]$，对应整体水平上的个体间差异性。

当我们使用 LSREL 且用模型 1 拟合表 2.2 的数据时，我们得到了表 2.3 中的参数估计值。[13]结果表明存在显著的、未得到解释的个体内方差($\hat{\theta}_\epsilon$)与个体间方差($\psi_{11}$)。在以上两种情况中，这些参数的显著性都可以从它们的大小相对于它们的标准误都超过了 2∶1 的比率看出来。显著的差异性为拟合更复杂的模型(如加入时变与非时变协变量)提供了依据。




注:CLSN=母亲—子女亲密度。

**图 2.2　一个表示随机截距模型的路径图**

同时，与图 2.1 暗示的结果一致，拟合度的参数也显示随机截距模型没有令人满意地拟合数据。卡方检验在 $p < 0.0001$ 的水平上拒绝了这一模型，RMSEA 的值远远超过了可以接受的范围(大约小于 0.08)，而且模型具有很大的残差。非常明显，一个只有截距的模型是不适合母亲—子女亲密度的数据的。接下来，一个线性截距因子被加入到模型中，以表示图 2.1 中观测到的大致的线性趋势。

表 2.3　模型 1:随机截距模型

| 参　数 | 估计值 |
| --- | --- |
| 平均截距 $\hat{\alpha}_1$ | 37.00(0.09) |
| 截距方差 $\hat{\psi}_{11}$ | 5.27(0.30) |
| 扰动项方差 $\hat{\theta}_\varepsilon$ | 4.68(0.11) |

注:括号中的数字是参数估计值的标准误。

## 模型 2:固定截距、固定斜率

在模型 2 中,我们把截距与斜率都固定,表示只估计一个唯一的、平均的截距参数($\hat{\alpha}_1$)与线性斜率参数($\hat{\alpha}_2$),并忽略任何变化方面的个体间差异性。我们通过在 $\mathbf{\Lambda}_y$ 中增加一列来包括斜率。这一列现在表示斜率因子并反映了数据的两个特性:第一,对应于一年级测度的负载被编码为 0,使时间原点定在一年级;第二,列中元素的间距反映了在二年级没有测度时的情况。因此,虽然第一个测度与第二个测度之间的时间是第二个测度与第三个测度之间时间的两倍,但由于时间编码的方式(见方程 2.5),模型仍然反映了线型的增长。

图 2.3 描述了模型 2 中设定的潜变量增长曲线。用矩阵表示截距与斜率因子,它们的方差、协方差以及平均值如下:

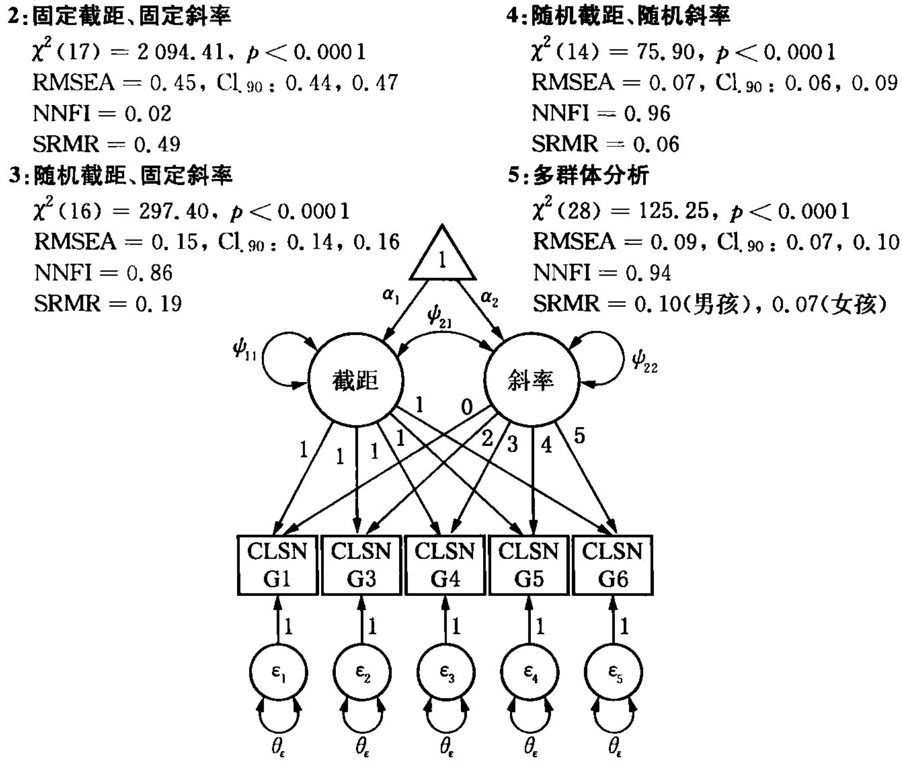
$$\mathbf{\Lambda}_y = \begin{bmatrix} 1 & 0 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \\ 1 & 5 \end{bmatrix} \qquad [2.5]$$

$$\mathbf{\Psi} = \begin{bmatrix} 0 & \\ 0 & 0 \end{bmatrix} \qquad [2.6]$$

$$\boldsymbol{\alpha} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} \qquad [2.7]$$

$$\boldsymbol{\Theta}_{\varepsilon} = \begin{bmatrix} \theta_{\varepsilon} & & & \\ 0 & \ddots & & \\ & & \ddots & \\ 0 & & 0 & \theta_{\varepsilon} \end{bmatrix} \qquad [2.8]$$

因为截距与斜率都是固定的，所以它们的方差与协方差都被限定为 0（因此 $\boldsymbol{\Psi}$ 包含的都是 0）。同模型 1 一样，矩阵 $\boldsymbol{\Theta}_{\varepsilon}$ 也被限定用来表示扰动项方差相同的假设。




注：CLSN = 母亲—子女亲密度。

**图 2.3 表示具有随机截距、随机斜率、截距—斜率协方差和相同扰动项方差的一般线性潜变量增长曲线模型的路径图**

表 2.4 与图 2.3 描述了模型的结果。新的斜率参数估计值 $\alpha_2$ 是显著的负值，反映了母亲与子女间的亲密度在小学阶段随时间的推移而降低的事实。但

是，模型 2 的拟合度（见图 2.3）比模型 1 的拟合度差很多。这一较差的拟合度来自截距与斜率因子设定的方式。模型 2 使用固定的截距与斜率。它限定每一对母亲—子女都有相同的起始亲密度，并强制每一对的亲密度都以恰好相同的速度降低。人与人之间显然是不同的，但这一模型并没有考虑到这一点。从另一个角度来看，这一拟合不足的原因在于，固定截距模型意味着方差的个体间部分（组内相关系数，即 ICC）为 0。数据中 ICC 偏离 0 的程度决定了固定截距模型拟合更差的程度。一个暗含非零 ICC 的随机截距模型能够解释重复测度间的部分自相关。我们先通过放松对截距方差的限制，或者说“解放”截距（模型 3）来解决这一问题，然后再同时解放截距与斜率（模型 4）。

**表 2.4　模型 2：固定截距、固定斜率**

| 参　数 | 估计值 |
|---|---|
| 平均截距 $\hat{\alpha}_1$ | 38.00(0.09) |
| 平均斜率 $\hat{\alpha}_2$ | −0.36(0.03) |
| 扰动项方差 $\hat{\theta}_\varepsilon$ | 9.57(0.21) |

注：括号中的数字是参数估计值的标准误。

## 模型 3：随机截距、固定斜率

我们有理由怀疑并不是每一对母亲—子女在一年级时有同等的亲密度。但是，先前的固定截距模型的设定却是这样假定的。虽然我们在拟合模型 1 时解放了截距，但是当我们把年级以斜率因子负载的形式引入后就又将其固定了。一个更实际的模型会通过允许截距的方差被自由地估计来允许个体在截距上有所不同。这一略微修改的模型通过估计一个平均截距（$\hat{\alpha}_1$）和一个表示个体的方差相对于总体平均方差的变化程度的截距方差（$\hat{\psi}_{11}$）来表示个体在截距上的差异性。模型 3 的参数估计值（见表 2.5）与模型 2 的相似，母亲与子女亲密度的平均变化大约是每年级−0.36 个单位而平均的截距大约是 38。截距的方差（$\hat{\psi}_{11}=5.37$）是显著的，显示个体在初始状态有明显的差异。允许个体的截距是随机的——只解放了模型 2 中的一个参数——显著地提高了模型的拟合度。

如我们所料,这一模型的拟合度比前面的固定截距模型要好得多,$\Delta\chi^2(1)=1797$。[14]但是,正如显著的扰动项方差($\hat{\theta}_\epsilon=4.21$)所示,模型仍然有提高的空间。正如有理由假设儿童在一年级的初始状态是不同的,我们同样有理由假设每一对母亲—子女的亲密度随时间推移的变化速度也是不同的。

**表 2.5 模型 3:随机截距、固定斜率**

| 参　数 | 估计值 |
|---|---|
| 平均截距 $\hat{\alpha}_1$ | 38.00(0.10) |
| 平均斜率 $\hat{\alpha}_2$ | −0.36(0.02) |
| 截距方差 $\hat{\psi}_{11}$ | 5.37(0.30) |
| 扰动项方差 $\hat{\theta}_\epsilon$ | 4.21(0.10) |

注:括号中的数字是参数估计值的标准误。

请注意,模型 3 的平均截距和斜率的估计值与模型 2 相同,我们所做的只是允许围绕平均截距的个体差异性存在。有趣的是,模型 3 中截距方差与扰动项方差之和等于模型 2 中的扰动项方差。在模型 2 中,截距的方差被限定为 0,迫使所有的个体间差异以扰动项方差的形式表现。在下一个模型中,截距和斜率都被允许有所变化。

## 模型 4:随机截距、随机斜率

到现在为止,我们所展示的都是如何拟合只有固定斜率的模型。但是,一般而言,个体间在基线水平和变化的速度上都会有差异。我们先前的每一个固定参数模型都忽略了部分个体的差异性。为了更好地反映数据的本质,我们现在设定截距与斜率方差参数都是随机的。在模型 4 中,每一个个体都被允许有不同的斜率和不同的截距。[15]这一模型除了在更多的情况下更为实际外,还允许对截距—斜率的协方差($\psi_{21}$)进行估计。在我们的例子中,截距—斜率的协方差指的是一年级时,母亲—子女亲密度与随时间推移的变化速度的相关程度。

同先前的模型一样,模型 4 的 LGM 的表述是较为简单的。为了设定这一

模型，我们解放截距与斜率因子的方差并添加一个截距—斜率协方差参数。矩阵 $\mathbf{\Psi}$ 因此变为：

$$\mathbf{\Psi}=\begin{bmatrix}\psi_{11} & \\ \psi_{21} & \psi_{22}\end{bmatrix} \qquad [2.9]$$

用 LISREL 估计模型 4 中的参数得到表 2.6 中的估计值。截距的方差相对于斜率的方差而言较大，而截距—斜率的协方差(0.25，对应于 0.40 的相关系数)是显著的。这表明有较高截距的儿童的负斜率较小，或者说在一年级时更高的母亲与子女的亲密程度的下降速度，比那些在一年级时不那么高的母亲与子女亲密程度的下降速度慢。

**表 2.6 模型 4：随机截距、随机斜率**

| 参 数 | 估计值 |
|---|---|
| 平均截距 $\hat{\alpha}_1$ | 38.00(0.08) |
| 平均斜率 $\hat{\alpha}_2$ | −0.36(0.02) |
| 截距方差 $\hat{\psi}_{11}$ | 2.98(0.29) |
| 斜率方差 $\hat{\psi}_{22}$ | 0.14(0.02) |
| 截距/斜率协方差 $\hat{\psi}_{21}$ | 0.25(0.06) |
| 扰动项方差 $\hat{\theta}_\epsilon$ | 3.70(0.10) |

注：括号中的数字是参数估计值的标准误。

卡方差异检验显示，这个具有随机截距和随机斜率的模型的拟合度比只有一个随机截距的模型 3 的拟合度有显著的提高，$\Delta\chi^2(2)=221.5$，$p<0.0001$。RMSEA、NNFI 与 SRMR 拟合指标同样都显示了好的拟合。如前所述，截距与斜率间的显著正协方差意味着在一年级时更亲密的母亲—子女往往较少经历亲密度的大幅下降。

在读者作出与其母亲有更亲密关系的儿童在剩余的小学学年中，亲密度只会较小地下降的结论之前，请回顾一下我们关于截距—斜率协方差($\psi_{21}$)解释的评论：必须注意的是，如果零时点被定在其他的年龄，截距—斜率协方差可能远没有这么令人印象深刻了。譬如，大部分被选中的儿童在上幼儿园时有相同的母亲—子女亲密度，当时间在一年级对中时，仅仅截距的不同就会导致我们观

测到的协方差。实际上，如果我们重新调整年龄变量，从而使截距定义在一年级之前的两年，我们就会得到一个－0.03的不显著的截距/斜率相关度。这突出了在拟合一个模型时决定将时间刻度的原点放在何处的重要性。

Mehta与West(2000)指出，如果线性LGM是数据的一个合适的模型，那么重复测度的真实分数的方差就会遵循一个二次的模式。用$t$表示时间，$t^*$表示时间原点，得到：

$$\sigma^2_{\xi(t)} = \psi_{11} + \psi_{22}(t - t^*)^2 + 2\psi_{21}(t - t^*) \qquad [2.10]$$

在这里，$\sigma^2_{\xi(t)}$是$t$时的真实分数方差，$\psi_{11}$、$\psi_{22}$和$\psi_{21}$分别是总体截距方差、斜率方差和截距—斜率协方差。换句话说，个体轨迹的集合就像一条蝶形领结或是扇形。在很多情况下，个体间的差异性在哪一点得到最小化(领结的"结")是我们感兴趣的。这一点被Hancock与Choi(2006)称为"孔径"，可简单地通过选择最小化截距方差的时间原点计算出来：

$$a = a^* - \frac{\hat{\psi}_{21}}{\hat{\psi}_{22}} \qquad [2.11]$$

在这里，$a^*$是最初选定的时间原点的时点，$\hat{\psi}_{21}$与$\hat{\psi}_{22}$分别是估计的截距—斜率协方差与斜率方差。[16]孔径是$\psi_{21} = 0$时的那一点(Mehta & West, 2000)。用模型4中的结果，儿童在母亲—子女亲密度上最相近的时点是$a = 0 - 0.2533/0.1366 = -1.85$，即进入小学的前一年。当然，当孔径落在观测到的数据的时点范围内时，我们就应当谨慎解释，因为这就迫使研究者假设线性模型适用于这些时点，而这并不一定成立。

总结我们的以上讨论，我们从估计一个空模型发展到估计一个有随机截距与随机斜率的模型。随机截距模型(模型1)表明，绝大多数的差异发生在个体内部，但个体之间也有不可忽视的差异性存在。模型2显示，当个体间差异被忽略时，我们可以发现什么，以及当不同的个体的结果测度有不同的变化轨迹时，将参数限定在某一特定值时会如何损害模型的拟合度。模型3基于模型2，解放了截距方差参数。我们发现，允许不同的人有不同的截距导致了模型拟合度的大幅提高。具有随机截距与斜率的模型(模型4)比先前限定更多的模型(模

型 2 与模型 3)表现得好得多。除了提高整体拟合度外，模型 4 的优势在于发现了截距与斜率间的一个显著的正协方差($\psi_{21}$)，这表明亲密度随时间推移而变化的速度与一年级时的亲密度有关。我们将模型 4 作为接下来研究的模型的基础。在模型 5 中，我们检验了母亲—子女亲密度变化轨迹存在性别差异的可能性。

## 模型 5：多群体分析

模型 4 的结果表明，假设个体具有不同的截距与斜率是有意义的。然而，这部分的差异有可能与其他变量有系统性的相关。例如，我们可以假设和检验男孩与女孩在平均截距与斜率上的群体差异(McArdle & Epstein, 1987)。至少有两种方法来检验群体在轨迹上的差异：(1)把样本分成两组，然后在两组中同时估计参数；(2)把群体变量设定为截距与斜率的自变量。在 LISREL 中，设定一个双群体分析需要将数据根据性别分为两个数据集，并进行一个多样本分析。在这一分析中，模型被同时拟合到两个数据集中。两个模型用相同的程序命令设定，每一个模型都如图 2.3 中所描述的那样。不同群体模型的参数矩阵中的重要参数被加上相等的限定，以检验各群体在这些参数上存在差异的假设。这一多群体分析的方法原则上可以被运用于任意数量的群体，而不同的群体也可以设定不同的模型形式。正如我们下一章要讨论的，这一方法使我们能够以一种新颖的方式来研究纵向设定下的实验效应以及初始状态与实验的交互效应。而用群体作为自变量的方法将在模型 6 中进行讨论。

我们把多群体的策略运用到亲密度数据上，对男孩组与女孩组同时拟合分别的模型，但并不设定跨群体的模型参数限制，其结果显示在表 2.7 中。其中，NNFI 是根据每个群体的双参数空模型分别计算的。模型拟合度一般(RMSEA = 0.088, 90%的置信区间 = {0.072, 0.104})，表明允许对参数估计值进行谨慎解释。男孩组与女孩组的所有参数估计值，除了女孩组的截距—斜率协方差($\psi_{21}$ =0.14)外都显著，这个值比男孩组的值($\psi_{21}$ = 0.36)明显低得多。与这些协方差所对应的相关系数分别是 $r$ = 0.22(女孩组)和 $r$ = 0.59(男孩组)，表明

在初始状态(一年级时)和随时间变化的母亲—子女亲密度的相关程度上,男孩比女孩要高得多。

表 2.7　模型 5:多群体分析

| 参　数 | 估计值 |
|---|---|
| 男孩 | |
| 平均截距 $\hat{\alpha}_1$ | 37.85(0.12) |
| 平均斜率 $\hat{\alpha}_2$ | −0.38(0.03) |
| 截距方差 $\hat{\psi}_{11}$ | 3.03(0.42) |
| 斜率方差 $\hat{\psi}_{22}$ | 0.12(0.03) |
| 截距/斜率协方差 $\hat{\psi}_{21}$ | 0.36(0.08) |
| 扰动项方差 $\hat{\theta}_\varepsilon$ | 3.78(0.15) |
| 女孩 | |
| 平均截距 $\hat{\alpha}_1$ | 38.14(0.11) |
| 平均斜率 $\hat{\alpha}_2$ | −0.33(0.03) |
| 截距方差 $\hat{\psi}_{11}$ | 2.90(0.39) |
| 斜率方差 $\hat{\psi}_{22}$ | 0.15(0.03) |
| 截距/斜率协方差 $\hat{\psi}_{21}$ | 0.14(0.08) |
| 扰动项方差 $\hat{\theta}_\varepsilon$ | 3.63(0.14) |

请注意,男孩组的截距与斜率(分别是 37.85 和−0.38)比女孩组的(分别是 38.14 和−0.33)斜率低。这表明平均而言,男孩似乎比女孩起点低,而下降的速度快。考虑到这些明显的差别,研究者有充足的理由来检验这些差别是否显著,一个方法就是限定这些跨群体的参数相等,然后通过卡方差异检验来观察模型的拟合度是否有显著的下降。在这里,如果我们限定截距的均值相等(允许斜率变化),差异检验的结果就是 $\Delta\chi^2(1)=3.11$, $p=0.08$,表明截距没有显著的不同。如果我们改为限定斜率的均值相等(允许截距变化),差异检验同样也是不显著的,即 $\Delta\chi^2(1)=1.47$, $p=0.23$,表明斜率也没有显著的不同。尽管看起来有区别,但并没有足够的证据表明男孩与女孩在一年级到六年级间的母亲—子女亲密度上遵循不同的线性轨迹。当然,检验其他跨群体限制条件同样也是可能的,但理论可能会建议检验扰动项方差或截距—斜率协方差($\psi_{21}$)的区别。关于男孩组与女孩组 $\psi_{21}$ 相等的假设,其检验也是不显著的,即 $\Delta\chi^2(1)=$

3.83，$p > 0.05$。

关于多群体 LGM 的例子和进一步的讨论，见 Curran、Harford 与 Muthén (1996)，Curran、Muthén 与 Harford(1998)，McArdle(1989)和 McArdle 与 Epstein(1987)的著作。在这一节结束时，我们有必要指出，我们的研究假设群体的成员资格是已知的(可观察到的)。如果群体是假定的，同时其成员资格是不确定的，那么研究者可能会对第 3 章讨论的增长混合模型感兴趣。接下来，我们将展示对于亲密度变化轨迹的性别差异研究的另一种方法。

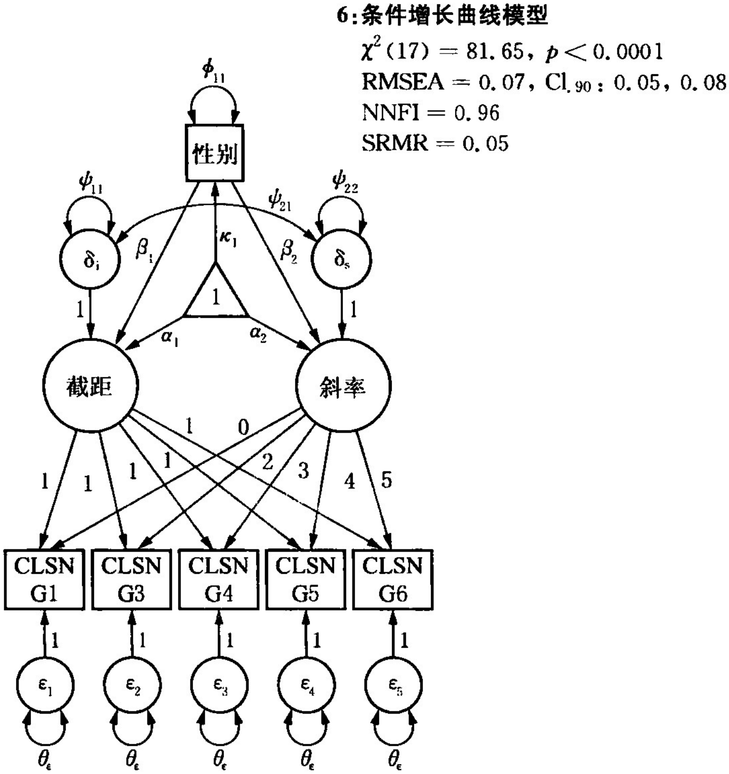
## 模型 6：条件增长曲线模型

除了使用多群体模型，对截距与斜率的性别差异分析可以通过在一个单组分析中，将性别作为截距与斜率的外生变量(第二层单位或非时变协变量)来大大简化模型。这样的随机变量的自变量能够被引入模型来解释个体间截距与斜率估计值的差异。方差参数在模型 3 与模型 4 中被解放。因为截距与斜率的方差可以被看做未得到解释的个体差异，这种分类或连续的第二层变量有可能解释这些差异。这类模型有时被称为“条件 LGM”(Tisak & Meredith，1990；Willett & Sayer，1994)，而模型 1 到模型 5 则可以被称为“无条件 LGM”(Singer & Willett，2003)。虽然我们只选择性别作为自变量，但实际应用中可以选择任意数量的自变量，如测量个体在种族、社会经济地位或父母的宗教信仰等方面的差异的变量。与模型 5 相比较，模型 6 的一个明显优势就是，因为不需要把样本分成几个群体，非时变协变量既可以是名义变量(如这里的性别)，也可以是连续变量。但其中一个不足就是，研究者必须能够假设某些跨群体的模型参数不变。譬如，我们在模型 5 中可以自由地估计男孩组和女孩组的不同扰动项方差，但在模型 6 中，我们就要投入许多额外的努力才能做到这一点。

我们在模型 6 中展示了运用分组变量(性别)作为自变量的方法。在这一模型中，我们用一个固定系数 $\beta_1$ 使性别成为截距的一个自变量，这一系数被解释为性别对截距的平均效果。[17]性别同样以固定系数 $\beta_2$ 为参数，作为斜率的一个自变量。当一个非时变协变量被当做斜率的自变量时，其效应常常被称为“跨

层交互效应”，因为时间（第一层）与协变量（第二层）互相影响，共同预测重复测度（Cronbach & Webb，1975；Curran，Bauer & Willoughby，2004；Kreft & de Leeuw，1998；Raudenbush & Bryk，2002）。注意，如果年龄对母亲—子女亲密度的影响在个体间是不同的，而性别部分地解释了个体在这一影响上的差异，那么跨层交互效应与传统的多元回归分析中的调节效果就有同样的解释。[18]

图 2.4 是一个包括性别作为截距与斜率的自变量的路径图。截距与斜率的方差以及它们的协方差被重新概念化为残差方差与残差协方差，即性别未能解释的那部分方差与协方差。在图 2.4 中，这些残差参数即 $\psi_{11}$、$\psi_{22}$ 与 $\psi_{21}$。参数 $\beta_1$ 和 $\beta_2$ 分别表示性别对截距和斜率的影响。这一模型的结果展示在图 2.4 和表 2.8 中。NNFI 是将模型 0 作为空模型计算出来的，因为估计性别的均值和方差得以提高。




注：CLSN = 母亲—子女亲密度。

**图 2.4　以性别自变量预测截距与斜率的个体差异的模型的路径图**

模型 6 的结果显示，性别预测截距的系数是−0.29，也就是说，在一年级时，母女(编码为 0)比母子(编码为 1)有较高的亲密度分数，但这一差异并不显著。母女也比母子表现出较少的随时间发生的变化，虽然也不显著。正如我们所预期的，这些结果同我们从模型 5 中得到的一致。例如，男孩和女孩的平均截距的差异(见表 2.7)是 37.85 − 38.14 = −0.29，等于模型 6 中所估计的性别的固定效应 $\hat{\beta}_1$(见表 2.8)。同样，模型 5 中的平均斜率差异是−0.05，等于模型 6 中得到的性别对斜率的固定效应 $\hat{\beta}_2$。由于模型 5 和模型 6 并不完全相同(跨群体间的扰动项方差不同)，结果不一定总是这么相近。我们须注意，这里的结果可以通过增加一个跨群体相等的限定而轻易改变。

**表 2.8　模型 6:条件增长曲线模型**

| 参　数 | 估计值 |
|---|---|
| 平均截距 $\hat{\alpha}_1$ | 38.14(0.11) |
| 平均斜率 $\hat{\alpha}_2$ | −0.33(0.03) |
| 性别对截距的影响 $\hat{\beta}_1$ | −0.29(0.16) |
| 性别对斜率的影响 $\hat{\beta}_2$ | −0.05(0.04) |
| 截距方差 $\hat{\psi}_{11}$ | 2.96(0.28) |
| 斜率方差 $\hat{\psi}_{22}$ | 0.14(0.02) |
| 截距/斜率协方差 $\hat{\psi}_{21}$ | 0.25(0.06) |

注:括号中的数字是参数估计值的标准误。

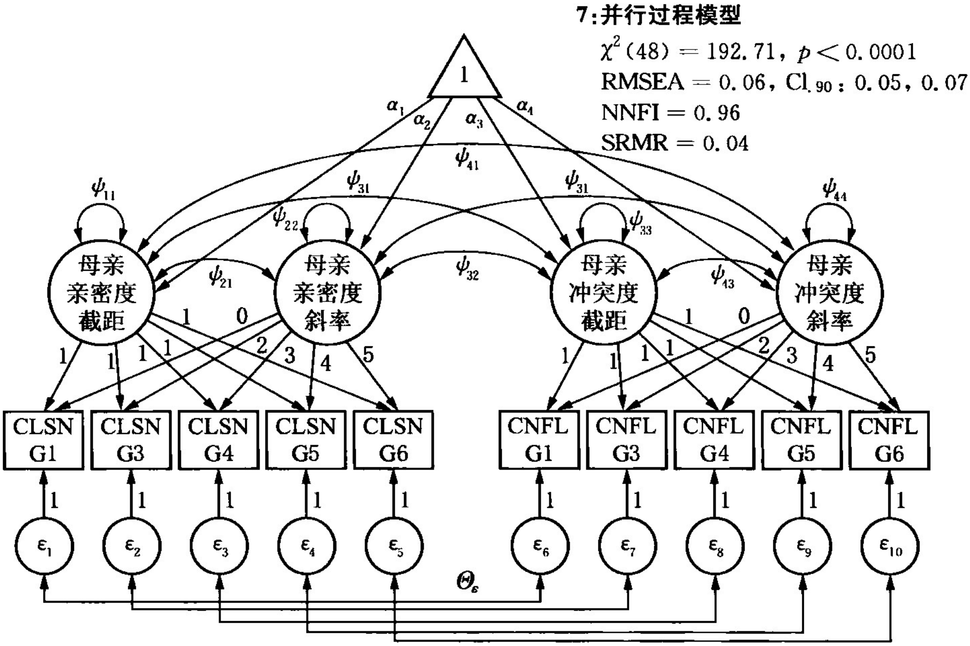
把非时变协变量当做增长因子的自变量可能存在问题。条件增长曲线模型可以被理解为一个调解模型。在这一模型中，增长因子被假定为完全调解非时变协变量对结果变量的影响，但完全调解不太可能是一个站得住脚的假设。例如，在模型 6 中，如果省略直接连接儿童性别与亲密度的路径，那么性别的直接影响就被默认为 0。如果这些零限定是不适当的，模型拟合度就会变得很差，参数估计值也可能是有偏的。一个可行的策略是放宽对直接影响的限定(从而估计直接连接协变量与每个重复测度的路径系数)而将 $\hat{\beta}_1$ 和 $\hat{\beta}_2$ 限定为 0(Stoel, van den Wittenboer & Hox, 2004)。但是，这个模型也许并不能从实质上解决问题。此外应当记住，外生变量影响截距因子的解释是会随着时间的测量尺度与时间原点的位置发生变化的(Stoel, 2003; Stoel & van den Wittenboer, 2003)。

## 模型 7：并行过程模型

我们可以研究两个重复测度的变量在变化方面的关系对于一个以上变量的增长过程进行拟合，这一过程允许我们研究不同变量变化方面的关系(McArdle, 1989)。譬如，研究者或许对同时拟合母亲—子女冲突度与父亲—子女冲突度的增长感兴趣，并探索一个增长曲线的截距与另一个斜率之间的关系。这种模型有“并行过程模型”(Cheong, MacKinnon & Khoo, 2003)、“多元变化模型”(MacCallum et al., 1997)、“跨域个体增长模型”(Sayer & Willett, 1998; Willett & Sayer, 1994, 1995)、“多域模型”(Byrne & Crombie, 2003)、“完全多元潜变量轨迹模型”(Curran & Hussong, 2003; Curran & Willoughby, 2003)、“同步增长模型”(Curranet al., 1996)、“二元增长模型”(Aber & McArdle, 1991)或“联合 LGM”(S. C. Duncan & Duncan, 1994; T. E. Duncan, Duncan & Strycker, 2006; T. E. Duncan et al., 1999; Tisak & Meredith, 1990)等各种名称，可以很容易从一个简单的随机截距、随机斜率模型中得出。并行过程模型包含两组截距与斜率，每个重复测度的变量各有一组。截距与斜率间的协方差同样被估计(见图 2.5)。在这个例子中，我们用母亲—子女亲密度与母亲—子女冲突度作为两个因变量(见表 2.9)。样本包括 849 个有母亲—子女亲密度与母亲—子女冲突度完整数据的儿童(433 个女孩和 416 个男孩)。NNFI 是将一个亲密度和冲突度的固定截距模型作为空模型计算的，允许特定时点的扰动项共变。

表 2.9 显示了一些有意思的结果，但对小学儿童的家长来说，也许没有什么是让人惊奇的。首先，如我们所料，与母亲—子女亲密度相关的参数估计值与模型 4 中的几乎相同，且任何差异都可以归结于模型 7 所用样本由于缺失数据的缘故而较小这一事实。亲密度与冲突度都随时间而变化，但是正如它们的平均斜率估计值所显示的，两者变化的方向相反。截距与斜率间协方差的估计值呈现在表 2.9“曲线协方差”部分，这些协方差暗含的相关度则在“曲线相关”的部分。截距的协方差($\psi_{31}$)是显著的负值，意味着与母亲在一年级时特别亲密的

儿童经历最少的冲突。同样，斜率协方差($\psi_{42}$)也是显著的负值，意味着那些亲密度急剧下降的儿童往往就是那些随着年龄增长而经历加速冲突的儿童。最后，冲突度的截距与亲密度的斜率($\psi_{32}$)是负相关的，说明那些在一年级时表现出与母亲相对较多的冲突的儿童，随着他们年纪的增长，往往会经历急速的亲密度的下降。




注：CLSN = 母亲—子女亲密度；CNFL = 母亲—子女冲突度。

**图 2.5 母亲—子女亲密度与母亲—子女冲突度同时增长模型的路径图**

**表 2.9 模型 7:并行过程模型**

| 参　　数 | 估计值 |
| --- | --- |
| 母亲—子女亲密度 | |
| 平均截距 $\hat{\alpha}_1$ | 38.00(0.08) |
| 平均斜率 $\hat{\alpha}_2$ | −0.36(0.02) |
| 截距方差 $\hat{\psi}_{11}$ | 2.99(0.29) |
| 斜率方差 $\hat{\psi}_{22}$ | 0.14(0.02) |
| 截距/斜率协方差 $\hat{\psi}_{21}$ | 0.25(0.06) |

续表

| 参　数 | 估计值 |
|---|---|
| 母亲—子女冲突度 | |
| 平均截距 $\hat{\alpha}_3$ | 15.24(0.20) |
| 平均斜率 $\hat{\alpha}_4$ | 0.30(0.04) |
| 截距方差 $\hat{\psi}_{33}$ | 25.74(1.61) |
| 斜率方差 $\hat{\psi}_{44}$ | 0.40(0.06) |
| 截距/斜率协方差 $\hat{\psi}_{43}$ | −0.57(0.23) |
| 曲线协方差 | |
| 截距协方差 $\hat{\psi}_{31}$ | −3.88(0.51) |
| CLSN 截距/CNFL 斜率协方差 $\hat{\psi}_{41}$ | 0.13(0.09) |
| CNFL 截距/CLSN 斜率协方差 $\hat{\psi}_{32}$ | −0.41(0.12) |
| 斜率协方差 $\hat{\psi}_{42}$ | −0.07(0.02) |
| 曲线相关 | |
| 截距协方差 | −0.44 |
| CLSN 截距/CNFL 斜率协方差 | 0.12 |
| CNFL 截距/CLSN 斜率协方差 | −0.22 |
| 斜率协方差 | −0.28 |

注：括号中的数字是参数估计值的标准误。
CLSN = 母亲—子女亲密度；CNFL = 母亲—子女冲突度。

拓展基本的并行过程模型是可能的。如果与两个变量的重复测度相关的斜率被假设成不仅相关，而且有因果关系的相关，那么就可以在增长因子间设定有方向性的路径。譬如，Curran、Stice 与 Chassin(1997)使用了一个并行过程模型，其中青少年酗酒与同辈人酗酒同时随时间发生线性变化。年龄、性别和父母酗酒被用来预测变化的方面，而个体在每个重复测度的变量截距(调整到最初的测度时点)上的差异被用来预测其他重复测度变量的差异性。Curran 和 Hussong(2002)对儿童的反社会行为与阅读能力进行并行增长建模，预测阅读能力的斜率与反社会行为的截距。Curran 等人(1996)设定了一个模型，其中酗酒和光顾酒吧的截距被假设影响彼此的斜率。Raudenbush、Brenan 与 Barnett(1995)用同样的方法为同时影响夫妻对婚姻质量评价变化的自变量建模，其中每一对夫妻中的每个人都以 3 年的间隔进行测量。此外，也可以设定超过两个重复测度变量的并行过程。

## 模型 8：队列序列设计

当在一个样本中评估变化轨迹时，横截面研究方法与纵向研究方法都有潜在的缺陷。一方面，横截面设计有时会存在队列或历史效应，这可能会导致研究者认为存在一个实际上并不存在的趋势，或者忽视一个真正存在的趋势。另一方面，纵向研究有时会受累于同一个体的重复测度的污染威胁。“队列序列设计”(Meredith & Tisak, 1990; Nesselroade & Baltes, 1979; Schaie, 1965、1986; Tisak & Meredith, 1990)，也称“加速纵向设计”(Miyazaki & Raudenbush, 2000; Raudenbush & Chan, 1992; Tonry, Ohlin & Farrington, 1991)或“收敛方法”(Bell, 1953、1954; McArdle, 1998)，通过结合纵向与横截面方法来研究发展变化，这一方法一直被建议用来减少这些潜在的混淆。队列序列设计同时也大大缩短了进行纵向研究所需的时间，并减少了样本损耗的问题(Tonry et al. , 1991)。考虑一下年龄作为时间度量的情况。与其追踪同一批高中一年级新生 8 年直到大学结束，研究者可以选择只追踪 3 个队列(高中一年级学生、高中三年级学生和大学一年级学生)4 年。通过运用研究对象的多个队列，同时对每个队列只进行几个时点的测量，就可以得到整个研究关注的时间段的完全轨迹。队列序列设计应当更多地被视为一种有效的数据收集策略而不是一种建模方法，但是这一策略带来了一些有意思的建模选择。

为了用我们的例子来展示队列序列数据的分析，我们创造了两个人为的队列。我们先把数据随机分为两组，然后删除部分数据来模仿一个从真正的队列序列设计收集来的数据模式。对于队列 1，我们删除了所有六年级儿童的测度，而对于队列 2，我们删除了所有一年级儿童的测度。这就生成了一个新的数据集，使得队列 1 的儿童有一年级、三年级、四年级和五年级的数据，而队列 2 的儿童有三年级至六年级的数据。在随机将个案分到每个队列并删除部分数据的过程中，我们的样本量减少到了 893 ($n_1 = 426$, $n_2 = 467$)。当然，在实践中，这两个队列的数据会是在同一个 5 年的时间段内同时收集的，甚至在时间尺度上更少的重叠也

是可以接受的。

队列序列数据可以用两种方法来研究。首先,研究者可以将数据合并,并进行一个单组分析。这一方法将每个队列因为没有收集数据而产生的缺失值当做完全随机缺失。这一般是一个保险的假设,因为缺失值是因研究设计而产生的(T. E. Duncan et al., 1999; B. Muthén, 2000)。或者,研究者可以把队列看成不同的组而进行一个多群体分析(McArdle & Hamagami, 1992)。这个方法与模型 5 相似,数据按队列分组,所有跨队列的相应参数都被限定为相等。这一多群体的选择来源于一种处理缺失数据的方法,其中相对较少的几个"缺失"模式是可识别的(Allison, 1987; T. E. Duncan et al., 1999; McArdle & Bell, 2000; McArdle & Hamagami, 1991; B. Muthen et al., 1987)。B. Muthén (2000)展示了如何用这两种方法来实现队列序列设计。应该注意的是,将不同的队列当做单独的群体时,可能会在某些情况下导致估计问题(T. E. Duncan et al., 1999)。对于某些样本量小的群体,甚至可能会出现比研究者所关心的时点更多的测度时点,这使我们不得不删除一些数据,即从模型中去掉一些测量时点。在这里,我们的样本量较大,所以这一问题并不重要。下面,我们展示单组和多群体两种方法。

单组方法的设定与模型 4 是一样的,因此路径图就是图 2.3 中的模型 4。它们唯一的区别就是所有队列 1 中的研究对象都缺失六年级的数据而队列 2 中的研究对象则都缺失一年级的数据。其结果(见表 2.10)与模型 4 的结果相似,

**表 2.10 模型 8:队列序列设计**

| 参数 | 单组估计值 | 多组估计值 |
|---|---|---|
| 平均截距 $\hat{\alpha}_1$ | 37.99(0.10) | 37.95(0.13) |
| 平均斜率 $\hat{\alpha}_2$ | −0.35(0.03) | −0.36(0.04) |
| 截距方差 $\hat{\psi}_{11}$ | 3.73(0.43) | 3.73(0.43) |
| 斜率方差 $\hat{\psi}_{22}$ | 0.10(0.03) | 0.10(0.03) |
| 截距/斜率协方差 $\hat{\psi}_{21}$ | 0.21(0.10) | 0.21(0.10) |
| 扰动项方差 $\hat{\theta}_\varepsilon$ | 3.77(0.12) | 3.78(0.12) |

注:括号中的数字是参数估计值的标准误。

虽然标准差较大，但考虑到较小的样本量和缺失数据，这也是预料之中的。尽管这样，结果支持从较大样本分析得来的同样结论。一年级平均亲密度分数的估计值 $\hat{\alpha}_1 = 37.99$，而平均斜率 $\hat{\alpha}_2 = -0.35$。两者都与模型 4 中的估计值相似，截距与斜率的方差和协方差也是一样的。

多群体队列序列方法与先前的双组分析(模型 5)相似，所有跨队列的参数，包括截距与斜率的均值，都被限定为相等。谨慎地设定这些相等的限定是非常重要的。在我们的分析中，每个数据集只包括那些参与者提供了数据的变量。因此，队列 1 的数据集只有 4 个对应于一年级、三年级、四年级和五年级的测度的变量。同样，队列 2 的数据集也只包含三年级至六年级的测度。队列 1 的四年级的亲密度变量是第三个重复测度，而在队列 2 中，则是第二个重复测度。两个队列的因子负载矩阵则是：

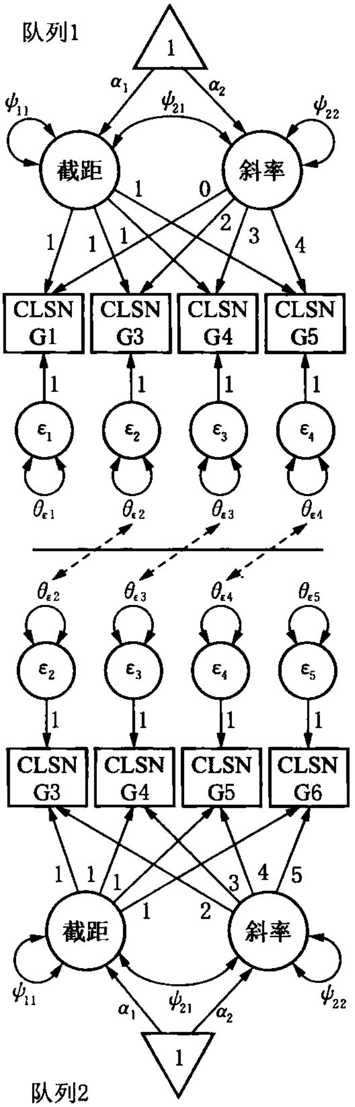
$$\boldsymbol{\Lambda}_1 = \begin{bmatrix} 1 & 0 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{bmatrix} \qquad [2.12]$$

$$\boldsymbol{\Lambda}_2 = \begin{bmatrix} 1 & 2 \\ 1 & 3 \\ 1 & 4 \\ 1 & 5 \end{bmatrix} \qquad [2.13]$$

与之对应的路径图显示在图 2.6 中。

表 2.10 中右边一栏的参数估计值与那些从单组分析中得到的估计值十分相似。关于多群体队列序列方法的另外的例子，见 Aber 与 McArdle(1991)，Anderson(1993)，Baer 与 Schmitz(2000)，Buist、Deković、Meeus 与 van Aken (2002)，T. E. Duncan、Dunca 与 Hops(1993)，S. C. Duncan、Duncan 与 Hops (1996)以及 AcArdle 与 Anderson(1990)的研究。

请注意，在单组队列序列设计中并没有提供 NNFI，这是因为部分数据被当做缺失。为了使用最大似然估计以获得以之为基础的所有拟合指标，我们建议




**8:队列序列设计**

单组

$\chi^2(14) = 73.16$, $p < 0.0001$

RMSEA = 0.07, $CI_{.90}$: 0.05, 0.08

多群体

$\chi^2(20) = 78.95$, $p < 0.0001$

RMSEA = 0.08, $CI_{.90}$: 0.06, 0.10

NNFI = 0.96

SRMR = 0.07(队列 1)

0.09(队列 2)

注:CLSN = 母亲—子女亲密度。

**图 2.6 多群体队列序列设计模型的路径图**

在可以获得每一组的完全数据的情况下使用多群体的方法,用协方差和平均值作为每一队列的输入数据。如果某些个案有缺失数据,我们建议采取完全信息最大似然估计法的缺失值估计技术以利用所有可得数据。放弃具有有效信息的数据在绝大多数情况下是不恰当的。此外,如果需要在每个队列内分别估计

扰动项方差，选多群体方法会比较好。

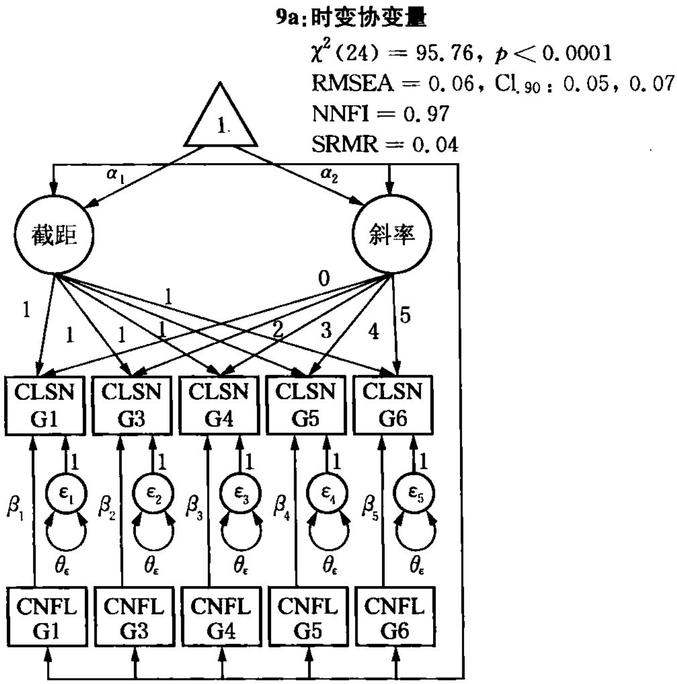
检验队列序列设计的多群体方法可以通过卡方检验以明确检验队列效应，如队列平均截距或斜率的差异(Anderson, 1993; Meredith & Tisak, 1990)。在单组的方法中，也可以通过将队列作为截距和斜率的一个虚拟自变量(Raudenbush & Chan, 1992)并通过其显著性水平来检验队列效应，但是这一设定需要有跨队列扰动项方差相等的假设。[19]

## 模型 9：时变协变量

前面我们把时变协变量(TVCs)定义为通过重复测量获得，并被用来预测一个结果变量的重复测度的变量。虽然有很多在多层模型的背景下关注TVCs的文章，但是它们很少关注如何在LGM的背景下处理TVCs。有两种方法可在LGM中处理TVCs。这两种方法讨论了略有不同的问题。第一种由B. Muthén(1993)提出并在George(2003)、Bijleveld与van der Kamp(1998)和Curran与其同事(Curran & Hussong, 2002、2003; Curran & Willoughby, 2003; Curran et al., 1998; B. O. Muthén & Curran, 1997)的文章中被阐明或被使用的方法，即直接将TVCs当做结果的重复外生自变量包括在模型中，如图2.7所示。这一模型中的$\beta$参数被解释为协变量的具体时点效应，或是协变量预测结果变量在具体时点上的偏差的能力。在这个方法中，一个TVC的效应可能会随时间而变化，但在个体内部是不变的。或者，$\beta$参数可以被限定为相等来表示协变量的效应在所有时点一直不变的假设。无论哪一种情况，这个模型都反映了重复测量变量在控制TVC具体时点效应后的增长。

我们用图2.7中的模型来拟合亲密度数据，把母亲—子女冲突度视为与亲密度同时测量的一个TVC。结果呈现在表2.11和图2.7中。NNFI是通过增强模型0(空模型)来计算的，也就是估计TVC的5个重复测度的均值和方差。冲突度对亲密度的具体时间效应在所有检验的年级保持在−0.15到−0.13之间，表明在每一时点去除个体差异由截距与斜率因子解释的部分后，冲突度常常与亲密度负相关。在没有满足建立因果关系的标准之前，将这一关系称为因

果关系是没有依据的。




注：CLSN = 母亲—子女亲密度；CNFL = 母亲—子女冲突度。

**图 2.7 以 B. Muthén(1993)建议的方式建立的包含时变协变量的增长曲线模型**

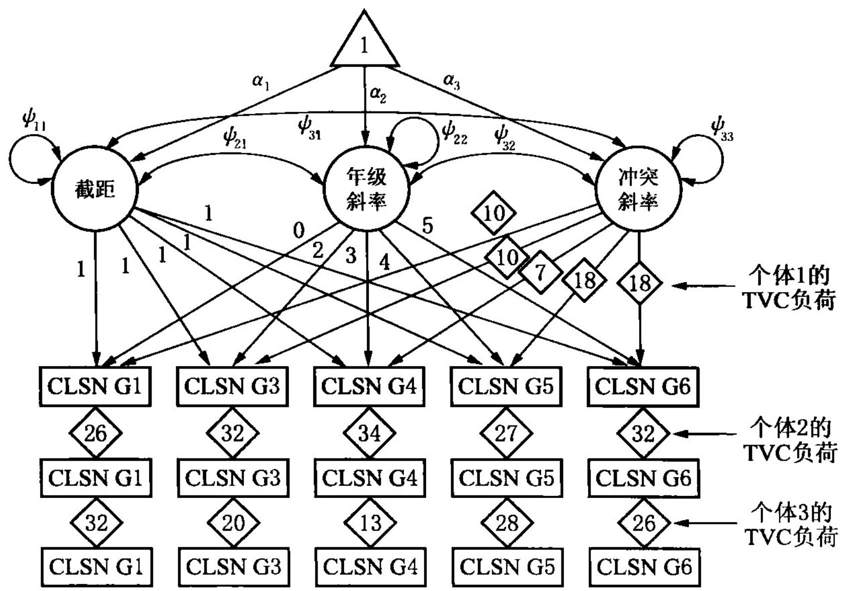
**表 2.11 模型 9a:时变协变量**

| 参　　数 | 估计值 |
|---|---|
| 平均截距 $\hat{\alpha}_1$ | 40.17(0.41) |
| 平均斜率 $\hat{\alpha}_2$ | −0.36(0.12) |
| 截距方差 $\hat{\psi}_{11}$ | 2.41(0.25) |
| 斜率方差 $\hat{\psi}_{22}$ | 0.13(0.02) |
| 截距/斜率协方差 $\hat{\psi}_{21}$ | 0.20(0.05) |
| 扰动项方差 $\hat{\theta}_{\varepsilon}$ | 3.52(0.10) |
| 冲突影响(G1) $\hat{\beta}_1$ | −0.15(0.03) |
| 冲突影响(G3) $\hat{\beta}_2$ | −0.13(0.01) |
| 冲突影响(G4) $\hat{\beta}_3$ | −0.13(0.01) |
| 冲突影响(G5) $\hat{\beta}_4$ | −0.13(0.01) |
| 冲突影响(G6) $\hat{\beta}_5$ | −0.13(0.02) |

注：括号中的数字是参数估计值的标准误。

另一个将 TVCs 纳入增长曲线模型的方法利用了定义变量。这个方法同 SEM 的标准做法有些不同,但与多层模型包含 TVCs 的标准方法(Raudenbush & Chan, 1993)类似。时间(或年龄等)变量本身可以被看做一个 TVC,因为时间在结果的重复测度上是不同的。如果定义变量能够容纳不同个体的不同测度时点,那么它们也能够以相似的方式被用来对其他 TVC 的效应建模。一个额外的斜率因子被添加到模型中来表示 TVC。TVC 的值被加入到个体数据矢量中,实质上给每个个体或第二层单位一个独特的 $\mathbf{\Lambda}_y$ 矩阵。在这一方法中,一个 TVC 的效应可以在个体间变化,但不随时间而变化。也就是说,被估计的是一个不同个体间的平均 TVC 效应,如果需要的话,TVC 的斜率因子的方差(以及同其他变化因子之间的斜方差)也可以被估计。这一方法要求将原始数据作为输入数据。

例如,来看图 2.8 中的模型。在每个测度时点(年级),每个个体也提供了母亲—子女冲突度(TVC)的数据。3 个个体数据矢量被作为例子,菱形框中的数




注:增长模型可以用于重复测量的协变量。扰动项方差被列出但没有解释。

CLSN = 母亲—子女亲密度;TVC = 时变协变量。

**图 2.8 以 Mehta 和 West(2000)建议的时间处理方式建立的包含时变协变量的增长曲线模型**

字表示每个 TVC 的个体数据矢量的内容。同样的标记法在 Mehta 与 West (2000:34)的著作中可以找到。使用这一方法，多个 TVCs 可以通过对每个协变量使用一个不同的斜率因子负载来纳入模型。TVCs 间的交互(如年级与母亲—子女冲突度之间的交互)可以通过将协变量的斜率负荷乘积作为斜率因子来研究。譬如，在图 2.8 中，个体 1 的交互因子的负载就是 0、20、21、72 和 80。非时变协变量与时变协变量间的跨层交互可以通过纳入 TVC 的斜率因子的预测值来设定。用这种方式设定的模型相当于带有第一层自变量的多层模型。

用图 2.8 中的模型拟合亲密度和冲突度数据，既而得到表 2.12 中的结果。因为是以原始数据作为输入数据，并未提供拟合指标(如 RMSEA 和 NNFI)。由于我们是用 Mx 来做这个分析，所以我们能够提供以似然值为基础的 95%置信区间而不是常用的标准误(Mx 并不提供标准误作为默认的选项，实际上，它警告不要对具有非正态样本分布的参数，如方差，使用标准误)。[20] 所有的参数估计值——除了两个之外——都是统计显著的。

**表 2.12 模型 9b:时变协变量**

| 参 数 | 估计值 | 95%置信区间 |
|---|---|---|
| 平均截距 $\hat{\alpha}_1$ | 40.12 | {39.82, 40.43} |
| 平均年级斜率 $\hat{\alpha}_2$ | −0.31 | {−0.35, −0.27} |
| 平均冲突斜率 $\hat{\alpha}_3$ | −0.14 | {−0.16, −0.12} |
| 截距方差 $\hat{\psi}_{11}$ | 4.93 | {3.40, 6.72} |
| 年级斜率方差 $\hat{\psi}_{22}$ | 0.11 | {0.08, 0.15} |
| 冲突斜率方差 $\hat{\psi}_{33}$ | 0.02 | {0.02, 0.03} |
| 截距/年级斜率协方差 $\hat{\psi}_{21}$ | 0.11 | {−0.08, 0.29} |
| 截距/冲突斜率协方差 $\hat{\psi}_{31}$ | −0.28 | {−0.39, −0.19} |
| 年级/冲突斜率协方差 $\hat{\psi}_{32}$ | 0.00 | {−0.01, 0.01} |
| 扰动项方差 $\hat{\theta}_\epsilon$ | 3.28 | {3.09, 3.48} |

## 模型 10：多项式增长曲线

到此为止，我们用增长曲线模型研究的轨迹都是时间或年龄的简单线性函

数,但 LGM 的使用者并不必局限于线性函数。迄今所提出的模型框架可以容纳任何在参数上是线性而在变量上是非线性的轨迹。也就是说,基本的 LGM 模型可以包容任何增长的参数来拟合与转化后的时间测量尺度相关的简单线性加权的轨迹。一个常见的例子是图 2.9 中的二次潜变量增长曲线(MacCallum et al. , 1997; Meredith & Tisak 1990; Stoolmiller, 1995)。在图 2.9 中,与二次斜率因子相关的负载是与线性斜率因子相关的负载的平方。二次斜率的均值($\alpha_3$)表示了轨迹中二次曲线的弯曲程度。

我们用图 2.9 中的二次增长曲线来拟合亲密度数据。结果呈现在表 2.13 和图 2.9 中。二次分子的均值并不显著($\alpha_3 = -0.019$, $p = 0.077$)。卡方差异检验显示,与单纯的线性模型相比,模型拟合度的提高是微不足道的,即 $\Delta\chi^2(1) = 3.12$, $p = 0.078$。如果有理论上的动机,我们可以选择允许二次斜率因子随机变化并与截距和线性斜率因子共变(Willett & Sayer, 1994),尽管并不存在一个固定而显著的平均二次项效应。实际上,这样做会导致模型拟合度的显著改善,但是很可能是因为额外增加的 3 个自由参数吸收了部分随机的差异,从而过度拟合了数据。

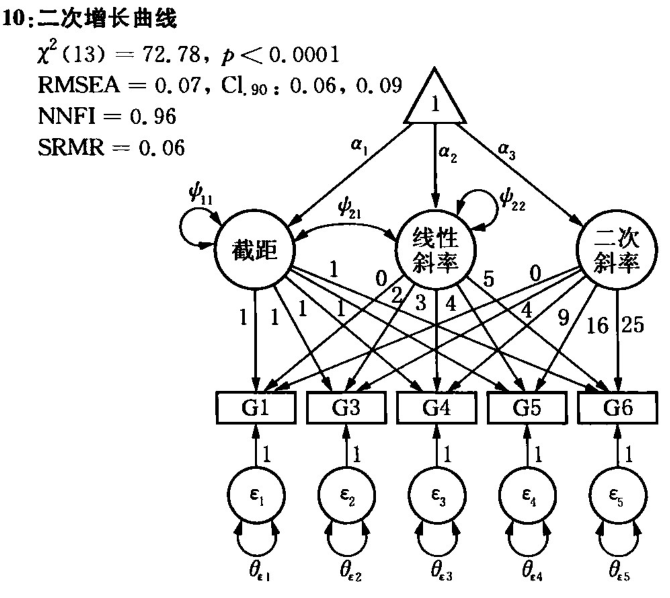



**图 2.9 一个包含固定二次斜率因子的潜变量增长曲线模型**

表 2.13 模型 10:多项式增长曲线

| 参　数 | 估计值 |
|---|---|
| 平均截距 $\hat{\alpha}_1$ | 37.94(0.09) |
| 平均线性斜率 $\hat{\alpha}_2$ | −0.26(0.06) |
| 平均二次斜率 $\hat{\alpha}_3$ | −0.02(0.01) |
| 截距方差 $\hat{\psi}_{11}$ | 2.98(0.29) |
| 二次斜率方差 $\hat{\psi}_{22}$ | 0.14(0.02) |
| 截距/线性协方差 $\hat{\psi}_{21}$ | 0.25(0.06) |
| 扰动项方差 $\hat{\theta}_\varepsilon$ | 3.70(0.10) |

注:括号中的数字是参数估计值的标准误。

我们并不认为在通常的情况下,当一个线性 LGM 不能充分拟合数据时,二次增长曲线就更有可能实现拟合。我们奉劝研究者不要使用以二次模型来描述某个特定的研究样本可能导致的独特性这种研究策略。我们怀疑,在社会科学中没有几个理论会自然地导致预测一个特定二次趋势的非线性变化。假定有足够的重复测度使额外的方差与协方差参数可以被识别,那么任意数量的多项增长因子都可以被加入到模型中。但是,这些做法必须基于一个适当的理论基础。

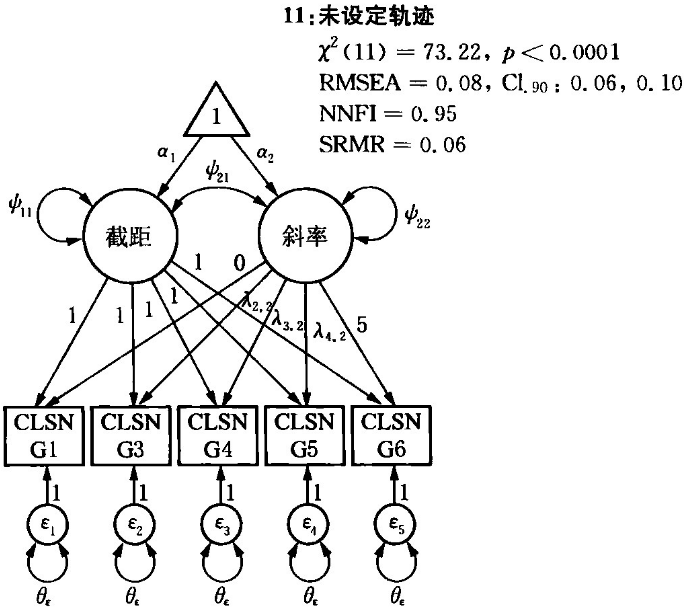
另外,我们也可以设定除了多项式曲线之外的函数形式。稍后我们会讨论结构潜变量曲线。这是对传统的增长曲线模型的拓展,它可以容纳更为复杂的、在参数上呈非线性的函数形式,其中增长的参数不再必须是简单的线性加权。例如,在对人口增长建模中常用的指数函数。

## 模型 11:未设定轨迹

到目前为止,在我们所描述的模型中,把结果变量和时间连接起来的函数是完全被定义的。例如,在图 2.3 中,从斜率因子到测量变量的路径在线性回归中被固定为从 0 到 5 的数值,对应于年级对母亲—子女亲密度的线性影响。关

于 LGM 的一个创造性的拓展涉及形态因子的建立：变化方面的增长函数的形态（及其因子负载）不是由研究者先验设定的，而是未知的且必须从数据中估计得来的（Meredith & Tisak, 1990）。例如，在模型 4 中，我们可以用形态因子替代线性因子，把这一新因子的第一个和最后一个负载分别限定为 0 和 1[21]，然后以 0 和 1 为基点，估计其他的 3 个负载（$\lambda_{22}$、$\lambda_{32}$和 $\lambda_{42}$）。被估计的负载将显示纵向数据的形态。虽然自由负载本身并不成比例，但是大致线性增长会导致负载从 0 到 1 单调递增。还可以将前两个负载限定为 0 和 1，在这种情况下，以发生在第一次和第二次之间的变化作为基准来解释接下来的间隔（Hancock & Lawrence, 2006; Stoel, 2003）。含有形态因子的模型有时候被称做"完全潜模型"（Curran & Hussong, 2002, 2003; McArdle, 1989）、"充分潜模型"（Aber & McArdle, 1991）或者"未指定模型"（T. E. Duncan et al. , 2006; Lawrence & Hancock, 1998; Stoolmiller, 1995; J. Wang, 2004），因为轨迹没有被先验地设定。这种模型比起前面讨论的模型具有更多的探索性意味，因为研究者不是在检验关于特定轨迹的假设。相反，数据被用来理解何种轨迹是适当的。解放某些线性斜率因子的负载来建立一个形态因子，允许通过嵌套模型卡方差检验——本质上是一个偏离线性的检验——对两个嵌套模型进行直接的比较。Duncan 等人（T. E. Duncan et al. , 1993），T. E. Duncan、Tildesley、Duncan 与 Hops（1995）以及 McArdle 和 Anderson（1990）提供了有关未设定轨迹模型的较好的例子。

我们把未设定轨迹模型用于拟合母亲—子女亲密度数据，将第一个和第五个负载分别固定为 0 和 5，使之与先前的时间标度一致（见图 2. 10）。比较模型 11 与模型 4 的卡方差异检验的结果是，它们没有显著的差异，即 $\Delta\chi^2(3) = 2.68$, $p = 0.44$。这表明一个线性的趋势足以对母亲—子女亲密度数据建模，因为斜率负载在没有限定的情况下仍然遵循一个近乎线性的趋势，如图 2. 11 所示。




注:CLSN = 母亲—子女亲密度。

**图 2.10 一个表示含有未设定形态因子的潜变量增长曲线的路径图**

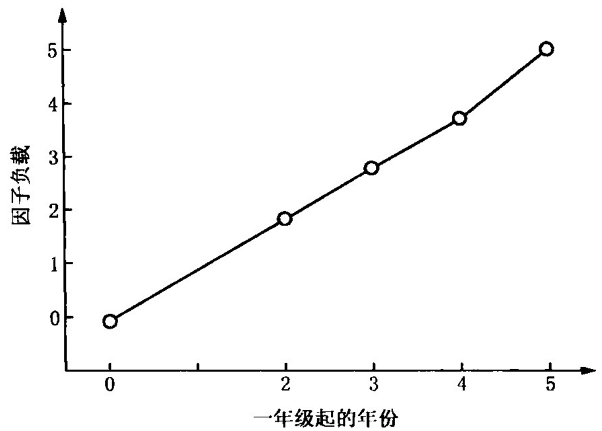



**图 2.11 从一年级起的斜率因子负载与年级的折线图**

## 总结

在这一章中,我们描述了几个潜变量增长曲线模型。从一个基本的空模型

开始，每个模型都被依次运用到同一个数据集来说明模型是如何实践运用的。除了带有随机截距和随机斜率的基本线性 LGM 外，我们展示了模型如何被拓展来处理多群体、纳入解释截距与斜率因子的自变量以及多于一个结果变量或一个年龄队列的增长的情况。我们还展示了如何将 TVCs 添加到一个增长曲线模型中以及如何为多项式或未设定的非线性轨迹建模。

我们要强调的是，研究者不需要被强迫来逐个研究这一章所描述的模型。如果理论或以前的研究表明一个随机截距、随机斜率的模型是恰当的，那么就没有必要拟合一个更为简单的模型。同样，如果有理由期待一个第二层自变量能解释个体在一个多项式趋势的二次因子上的差异，那么将模型 6 与模型 10 结合就是理所应当且恰当的，而不是以其中的一个或另一个作为开始。简言之，重要的是我们总要首先依赖于理论。如果理论不够明确，无法建议应该研究的模型，那么本章所展示的探索性策略——从一个空模型出发，逐步包含线性斜率因子，并添加随机效应与自变量——对于帮助研究者理解数据并恰当地对随时间而变化的变量建模是有用的。

# 第 3 章 | 模型的特别拓展

讨论至此，我们并没有穷尽潜变量增长模型的所有可能性。随着科学检验的日益复杂，过去检验这些假设的模型也因此在复杂性和灵活性上有所增加。在本章中，我们会利用最近一些统计理论的发展和软件技术的提高，呈现几种有趣且比较有用的拓展和一些传统增长曲线模型的特殊应用。这些应用包括增长混合模型、分段增长曲线模型、潜变量变化模型、结构潜变量曲线模型、自回归潜轨迹模型和对分类变量变化的建模。

## 增长混合模型

在前面讨论的大部分模型中，我们假定单个的潜变量轨迹会充分反映总体中的变化模式，同时又允许曲线有随机误差。在模型 5 中，我们注意到离散型变量(例如，男性和女性)可以有不同的变化曲线。也就是说，不同的增长曲线能够在几个已知的组群中被同时估计，并且关键参数的估计有无跨组的限制皆可。然而，分类变量并不一定要直接测量，潜在的(不可观测的)类型也有可能产生异质曲线。如果这样的曲线存在一个以上，而仅仅设定一个轨迹就可能产生显著的偏差，并得出错误地代表所有轨迹类型的轨迹(Sterba, Prinstein & Cox, 2007; Von Eye & Bergman, 2003)。

如果可以合理地假定异质曲线有潜在的来源，研究者就可以应用潜变量增长混合模型(LGMM)。在一个潜变量增长曲线混合模型中，假定总体包含 $K$ 个同质小组的混合，每个小组有它自身的变化轨迹。有两种通用的方法来拟合增长曲线混合模型(B. Muthén, 2001; Nagin, 1999; Nagin & Tremblay, 2001),

这两种方法都包括在一个潜在的分类变量上计算的回归潜截距和斜率，并且都允许轨迹的形式在不同类型中不同(例如，轨迹在一个类型里可以是线性变化，在另一个类型里可以是二次方程变化)。两种方法之间的主要不同之处在于，Muthén(2001)的方法允许轨迹在同一个等级内有变化，而Nagin(1999)的方法要求轨迹在不同等级之间变化。因此，较之前者，Nagin(1999)的方法通常会得出更多的潜在类型存在的结论。

LGMM的一个潜在的缺点是，这一模型可能致使研究者相信多个同质小组的存在，而实际上仅仅存在一个这样的小组，其数据分布是非正态分布或服从一个非线性的趋势(Bauer & Curran, 2003a)。换句话说，增长混合模型识别的组也许不代表真正的组，而是一个混合轨迹分布的组成部分，将这些组成部分放在一起，大致像一个单个的非正态分布。因为异质次总体的存在是潜在增长混合模型的一个基本假定，而这个方法不能证明有$K$个类别的存在，就像传统的潜变量增长曲线模型并未证明仅隐含一个总体曲线一样(Bauer & Curran, 2003a、2003b、2004)。一般来说，有大量的证据可以证明LGMM的很多假定被违反，并产生了不正确的潜在类型(Bauer, 2005、2007)。

增长混合模型不太容易被理解和操作。LGMM经常会产生错误的结果或者扩大分类，其时间起始值也经常会影响参数估计。除此之外，模型评估和模型选择也不是很直接。虽然设定不同分类数目的模型可以用贝叶斯信息准则来比较，但是这些标准在混合模型中的应用仍然有很多争议。在各个阶段，都需要作出几个比较主观的判断。再次，混合模型比起标准的模型常常需要更大的样本量。最后，虽然在潜变量分析中违反假定条件总是一个潜在的危险，但是潜变量增长混合模型是最容易受影响的。违反分布假设和错误设定曲线的形式可能导致把同质总体分成多重类别。即使缺点重重，但这些方法仍然可能实现令人满意的应用。例如，增长混合曲线模型可以检验包含等效原则的发展理论，即不同的起始条件会产生相同的结果，以及多瓣发展理论，即相同的起始条件会产生不同的结果(Cicchetti & Rogosch, 1996)。

混合增长曲线模型是一个正在被积极研究的领域，不断有新的研究成果出现。例如，Klein和Muthén(2006)描述了一个混合增长曲线模型的拓展，允许

增长的异质性依赖于初始状态和非时变协变量。较之标准的增长曲线模型，这种方法会产生更准确的间隔预测，但是这种方法在大部分软件中还没有实现。

目前，增长混合模型能够通过 MECOSA 中的高斯程序(Arminger, Wittenberg & Schepers, 1996)、Mplus(L. K. Muthén & Muthén, 1998—2006)、Mx(Meale et al., 2003)以及 SAS PROC TRAG(Jones, Nagin & Roeder, 2001)实现。目前，Mplus 是这些软件中最灵活的。有兴趣的研究者可以参考 Bauer(2005、2007)，Bauer 与 Curran(2003a、2003b、2004)，T. E. Duncan 等人(2006)，Li，Duncan、Duncan 和 Acock(2001)以及 M. Wang 与 Bodner(2007)的著作。

## 分阶段增长

假设以前的理论或者研究显示，增长应该是以不同的速度进行的，但仍然是线性的。例如，相比小学阶段而言，在中学阶段可能不同(也许是因为不同学校的资金来源不同)。两个增长阶段可以在单一的增长曲线模型中用分阶段增长模型(T. E. Duncan et al., 2006; Sayer & Willett, 1998)或者间断点设计(Hancock & Lawrence, 2006)来建模。通过在一个模型中纳入两个或者更多的线性斜率因子，可以设定分阶段增长模型。例如，如果我们连续在七年级至九年级收集数据，并假设在中学时增长速度会下降，那么负载矩阵 $\mathbf{\Lambda}_y$ 看上去会像方程 3.1：

$$\mathbf{\Lambda}_y = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 2 & 0 \\ 1 & 3 & 0 \\ 1 & 4 & 0 \\ 1 & 5 & 0 \\ 1 & 5 & 1 \\ 1 & 5 & 2 \\ 1 & 5 & 3 \end{bmatrix} \qquad [3.1]$$

在方程 3.1 中，第一列代表截距，第二列代表直到六年级为止的线性增长（第五次测量），第一个线性斜率因子的平均值代表线性增长的速度，这个速度是描述小学增长阶段的。矩阵 $\mathbf{\Lambda}_y$ 的第三列代表中学阶段的线性增长，把小学阶段的最后一年看做第二个阶段的起始点。第二个阶段斜率的平均值看做第二个阶段的增长速度。另外一种设定（但在统计上是相当的）可以用方程 3.2 中的 $\mathbf{\Lambda}_y$ 矩阵（Hancock & Lawrence, 2006）。在方程 3.2 中，第一个斜率因子代表从一年级到九年级的线性增长，第二个斜率因子代表任何其他的线性变化，七年级之后的变化可被这一线性斜率捕获。

$$\mathbf{\Lambda}_y = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 2 & 0 \\ 1 & 3 & 0 \\ 1 & 4 & 0 \\ 1 & 5 & 0 \\ 1 & 6 & 1 \\ 1 & 7 & 2 \\ 1 & 8 & 3 \end{bmatrix} \qquad [3.2]$$

最近，一些方法研究的成果展现了一些激动人心的可能性，那就是对发生在离散时段的变化建立模型。Cudeck 和 Klebe（2002）展示了长期纵向数据的多阶段建模，每个阶段可以有不同的函数形式（例如，一个下降的斜率会紧跟一个向上的斜率）。从一个阶段到下一个阶段的转折点被叫做“变化点”或者“结”。从理论上来说，在建模时，这些变化点有自身变化的方面——比如通过固定参数，待估参数或者随机系数来设定。例如，在人的一生中，非语言智力的发展变化中，Cudeck 和 Klebe 对此建立急速的二次增长，并逐渐变成线性下降。他们将年龄从第一阶段到第二阶段转变的评估作为一个随机系数，以 18.5 为平均数，第二层的方差是 9.25。

不是任何时候都能够在增长曲线模型框架中设置多阶段模型，但是如何检查模型拟合度、用变化特征作为预测变量、评价潜变量中心多阶段变化都值得

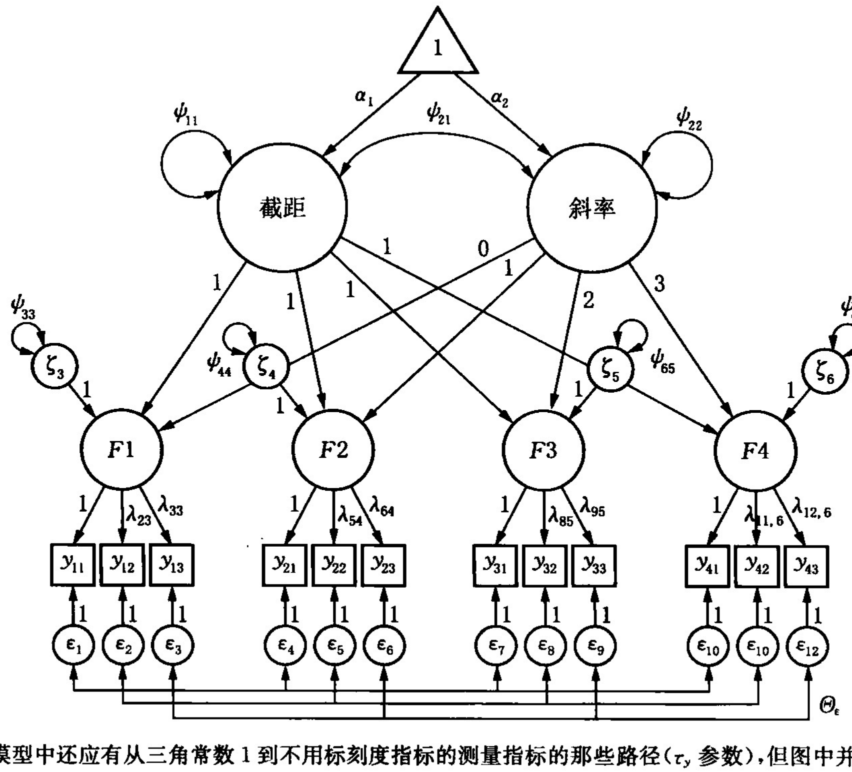
进一步探讨。如果已知变化点,可以用分割矩阵 $\mathbf{\Lambda}_y$ 来检测多阶段模型,每个分区的基础包含曲线轨迹的相应部分。例如,在方程 3.3 中,$\mathbf{\Lambda}_y$ 矩阵代表在等距间隔的情况下,在第一阶段有一个截距和线性斜率(均值分别为 $\alpha_1$ 和 $\alpha_2$),而在第二阶段有一个不同的截距和线性斜率(均值分别为 $\alpha_3$ 和 $\alpha_4$),两个阶段之间有一个变化点,时间等于 2.3。另外,如果按照方程 3.4 那样设定 $\mathbf{\Lambda}_y$ 矩阵,一个未知的变化点 $\omega$ 也可以被估计,并将它作为模型参数的一个函数。在这个模型中,时间测量尺度从第二列中脱离,但是会在第四列重新进入模型。假定分段 $(\alpha_1 + \alpha_2 time)$ 和 $(\alpha_3 + \alpha_4 time)$ 在时点 $time = \omega$ 时是连续的,那么 $y$ 值所表示的数值在多阶段曲线的两个阶段必须在 $\omega$ 时相等。因此其中一个参数(例如,$\alpha_3$)是多余的,可以通过限制其等于其他参数的函数来取消对其的估计,即设定 $\alpha_3 = \alpha_1 + (\alpha_2 - \alpha_4)\omega$,而将 $\omega$ 作为一个参数估计。对于更复杂的多阶段模型和更复杂的多阶段增长函数,相应的数学模型可以被应用到增长模型中。目前没有简单的办法对随机系数变化建模,但是此问题可以在多层模型中解决(Cudeck & Klebe, 2002)。

$$\mathbf{\Lambda}_y = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 2 & 0 & 0 \\ 0 & 0 & 1 & 0.7 \\ 0 & 0 & 1 & 1.7 \\ 0 & 0 & 1 & 2.7 \end{bmatrix} \qquad [3.3]$$

$$\mathbf{\Lambda}_y = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 2 & 0 & 0 \\ 0 & 0 & 1 & 3 \\ 0 & 0 & 1 & 4 \\ 0 & 0 & 1 & 5 \end{bmatrix} \qquad [3.4]$$

## 多指标潜变量的变化模型

前面提到的模型都没有充分利用结构方程模型最基础且最有用的特征：对多重测量指标的潜变量之间关系进行建模的能力。我们尚未讨论过任何可靠性和测量误差的问题，但是实际上在社会科学中，几乎不可能有接近完美的可靠性。通常结构方程模型可以把可测量的方差分为共同方差（一组可测量的变量的共同方差）和特别方差（特定的变量有特定的测量误差和方差可靠性）来为不可靠性建模。在大多数结构方程模型的应用中，$\Theta_\varepsilon$ 对角线的部分代表特别方差，但在潜变量增长模型中，$\Theta_\varepsilon$ 对角线代表测量误差和偏离均值程度的总和。




注：模型中还应有从三角常数 1 到不用标刻度指标的测量指标的那些路径（$\tau_y$ 参数），但图中并未展现（Hancock et al.，2001）。由于并不指望在不同时点测量的指标之间的协方差可以完全用它们对增长因子的共同依赖来解释，我们允许相似指标的独特因子在各个重复测量之间具有共变性。

**图 3.1 有 3 个测量指标的潜变量的 4 个重复测度的线性变化的二阶增长曲线模型**

为了把共同方差和特别方差区别开来，重复测量可以用潜变量来代表而多重时点标识变量可以直接引入模型中。这些模型有时候被叫做“曲线因素模型”(S. C. Duncan & Duncan，1996；McArdle，1988)、“潜变量纵向曲线模型”(Tisak & Meredith，1990)或者“二阶潜变量增长曲线”(Hancock，Kuo & Lawrence，2001；Hancock & Lawrence，2006；Sayer & Cumsille，2001)。重复测量的潜变量被称为“一阶因子”，而增长因子(如截距和斜率)被称为“二阶因子”。请参看图 3.1 中的一个例子。

当有多指标测量时，用潜变量重复测量模型有几个优点。第一，二阶潜变量增长曲线模型能够在重复测量中清楚地辨认测量误差的出现，并利用潜变量建模，调整误差的出现。第二，二阶增长曲线模型允许区分由于偏离均值而导致的干扰项差异(临时不稳定性，在图 3.1 中由 $\psi_{33}$ 至 $\psi_{66}$ 反映)和由测量误差导致的特别差异(不稳定性，在图 3.1 中由 $\Theta_\varepsilon$ 反映)。第三，二阶增长曲线模型可以检测纵向因子的不变性或者静止性(Sayer & Cumsille，2001；Tisak & Meredith，1990)。对于关注的潜变量，保证其在整个测量周期中含义不变是非常重要的(Willett，1989)。要想让这个假设成立，在整个重复测量的过程中，因子结构是不应该变化的(Chan，1989；Meredith & Horn，2001)。也就是说，至少对于相似的项目，因子负载在重复测量上应当是一样的。虽然探究纵向因子不变之类的问题超出了本书的范畴，但无论如何强调这类问题对理解增长随时间变化的重要性都不为过。关于设定二阶增长曲线的更多细节，请参考 Han (1998)、Hancock 等人(2001)以及 Sayer 与 Cumsille(2001)的著作。

## 结构潜变量曲线

早期的多项增长函数具有一个特点，那就是“动态一致性”。大致说来，动态一致性指的是“曲线平均”的特质，它与“平均曲线”有同样的函数形式(Singer & Willett，2003)。这个特征适合线性增长、二次增长模型和任何包含加权的并随时间变化的线性变化的增长函数。至于增长参数(如果以传统的多项式形式表达)，严格意义上动态一致性的结果应该是增长函数的一阶导数为简单的数字，这些数字能够直接在 $\Lambda_y$ 矩阵中编码，并在实际的结构方程模型的软件中应用。

但如 Merdith 和 Tisak(1990)所建议的，潜变量增长曲线模型的增长函数不一定被限定为多项式曲线。

Browne 与 du Toit(1991)以及 Browne(1993)著作中曾提出并说明了结构潜变量曲线的方法以建构没有动态一致性的非线性增长函数模型。在结构潜变量曲线中，$\mathbf{\Lambda}_y$ 的负载可以与任何假定的增长函数 $f(t, \theta)$(时间 $t$、增长参数为 $\theta$ 的函数)在数值上保持一致，这一函数被称为"目标函数"(Blozis, 2004)。函数 $f(t, \theta)$被认为是平滑的，并随着 $\theta$ 的不同而不同。结构潜变量曲线方法设定的增长模型是 Rao(1958)的 EFA(Exploratory Factor Analysis)方法所获得的增长函数参数的一种直接的拓展。矩阵 $\mathbf{\Lambda}_y$ 的元素不是设定的固定值，它们是被估计的，但要求与基础的曲线函数 $f(t, \theta)$保持一致。前面例子中提到的多项式曲线模型是这个更普遍的框架中的特殊例子。

为了更好地理解结构潜变量曲线，对于每一个增长参数，其在矩阵 $\mathbf{\Lambda}_y$ 中为多项增长曲线模型设定的负载对应于假定的增长函数的一阶偏导数。例如，在模型 10 中设定的二次增长曲线中，目标函数如方程 3.5 所示：

$$\hat{y}_{it} = \theta_1 + \theta_2 t_{it} + \theta_3 t_{it}^2, \; t = \{0, 2, 3, 4, 5\} \qquad [3.5]$$

对于每一个增长参数，这个函数的一阶导数各自是：

$$\frac{\partial \hat{y}_{it}}{\partial \theta_1} = 1 \qquad [3.6]$$

$$\frac{\partial \hat{y}_{it}}{\partial \theta_2} = t \qquad [3.7]$$

和

$$\frac{\partial \hat{y}_{it}}{\partial \theta_3} = t^2 \qquad [3.8]$$

这些是已知量。因此，

$$\mathbf{\Lambda}_y = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \\ 1 & 4 & 16 \\ 1 & 5 & 25 \end{bmatrix} \qquad [3.9]$$

更复杂的但动态不一致的增长函数(例如,指数增长、Gompertz 函数和 logistic 曲线)可以用类似的方法设定。对于动态不一致的增长函数,增长参数 $\theta$ 不能被简化成 t 的简单函数,它可能需要更复杂的 $\mathbf{\Lambda}_y$ 设定。例如,对于指数增长模型,如果一个指数变化适于拟合父母与子女关系的变化(事实上并不是,我们只是假定),目标函数将会是:

$$\hat{y}_{it} = \theta_1 - (\theta_1 - \theta_2)e^{(1-t_i)\theta_3},\ t = \{1, 3, 4, 5, 6\} \quad [3.10]$$

对于每一个参数,这个函数的一阶导数各自是:

$$\frac{\partial \hat{y}_{it}}{\partial \theta_1} = 1 - e^{\theta_3(1-t_i)} \quad [3.11]$$

$$\frac{\partial \hat{y}_{it}}{\partial \theta_2} = e^{\theta_3(1-t_i)} \quad [3.12]$$

和

$$\frac{\partial \hat{y}_{it}}{\partial \theta_3} = (\theta_1 - \theta_2)(t_{it} - 1)e^{\theta_3(1-t_i)} \quad [3.13]$$

因此,如果我们让最初的值 $t_{it}=1$,则有:

$$\Lambda_y = \begin{bmatrix} 0 & 1 & 0 \\ 1-e^{-2\theta_3} & e^{-2\theta_3} & 2(\theta_1-\theta_2)e^{-2\theta_3} \\ 1-e^{-3\theta_3} & e^{-3\theta_3} & 3(\theta_1-\theta_2)e^{-3\theta_3} \\ 1-e^{-4\theta_3} & e^{-4\theta_3} & 4(\theta_1-\theta_2)e^{-4\theta_3} \\ 1-e^{-5\theta_3} & e^{-5\theta_3} & 5(\theta_1-\theta_2)e^{-5\theta_3} \end{bmatrix} \quad [3.14]$$

对于多项式曲线模型,$\theta_1$、$\theta_2$ 和 $\theta_3$ 是 3 个基础曲线因子的均值估计。但在指数的结构潜变量曲线中,$\theta_1$ 和 $\theta_2$ 是估计均值,而 $\theta_3$ 是用 LISREL 的附加参数特性来估计的。这类模型的详细讨论在 Browne(1993)和 Blozis(2004)的著作中有所描述。Blozis 在其 2006 年的研究中(Blozis, 2006)拓展了这种方法来为多指数测量的潜变量建构非线性模型,并在 2007 年的研究中将该方法延伸至探索多变量非线性变化(Blozis, 2007)。我们需要专门的软件(du Toit & Browne, 1992)来估计大多数结构潜变量曲线,但是 LISREL 通过复杂的等式

限定是可以用来估计这样的模型的。

## 自回归潜变量曲线模型

虽然在研究随时间变化而发生的变化时，潜变量增长曲线模型是一个有用且灵活的方法，但是还存在其他以结构方程模型为基础的方法。其中一个方法是 Guttman 提出的自回归或者马尔科夫单一模型（Guttman, 1954），其中每一个重复的测量被看做是一个以前的测量和在某一具体时点上的干扰项的函数。与传统的潜变量增长曲线模型不同，这种单一模型不考虑随时间变化的均值结构，但是会用以前时段收集的测量数据来解释方差的变化。Curran 与 Bollen(2001)以及 Bollen 与 Curran 的著作(2004、2006)提出自回归潜变量轨迹模型(ALT)，这种模型既包含潜变量增长模型，也包括单一模型。一个自回归潜变量轨迹模型的例子在图 3.2 中有所描述。ALT 模型主要在两个方面不同于标准的潜变量增长曲线模型：第一，在图中所描述的 ALT 模型参数化过程中，没有与第一个重复测量相关联的干扰项；第二，像自回归和单一模型一样，直接路径（图 3.2 中的 $\rho$ 参数）被设定以连接相邻的重复测量。

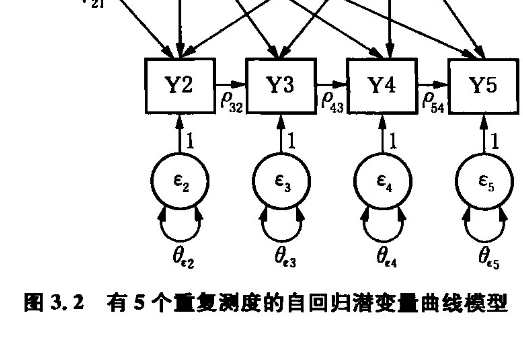



图 3.2　有 5 个重复测度的自回归潜变量曲线模型

通常情况下，单一模型和 ALT 模型会限定 $\rho_{t,t-1}$ 参数相等，虽然这样的限定并不是必需的。ALT 模型和之前提过的一个 TVC 模型有关联（模型 9，见表 2.8），其中在时点 $t-1$ 的结果变量被用来当做 $t$ 时点的一个自变量。单个变量或多元变量在 ALT 模型中的应用可以在 Rodebaugh、Carran 与 Chambless(2002) 以及 Hussong、Hicks、Levy 和 Curran(2001) 的著作中找到。一个与之更相关的模型是潜变量差异分数模型（McArdle, 2001; McArdle & Hamagami, 2001）。

## 分类变量和次序变量模型

至此，我们所讨论都是基于重复测量的变量是定距或定比性质的假设。然而，在社会科学中，许多变量常常是定序变量。例如，二分变量或者只有 3 个或 4 个选择的 Likert 量表就很难被假定具有定距或定比变量的特征（虽然在有很多选择的时候，它们可能被看做是连续的）。当重复测量的变量是定序变量时，这个数据包含的就不是均值或者协方差，而是一个潜在的大型列联表，在此之中，每一个单元格包含的是应答者所对应的一个特别的应答模式。如果这个定序数据能够被认为反映了一个潜在的正态分布变量，一个多阶段的估计过程就可以用来对增长建模（Mehta, Neale & Flay, 2004）。这个过程假设在每一次测量时，潜变量分布中观测不到的阈值会决定与应答模式关联的多元概率。在第一阶段，设定一个关联函数来对多元定序应答概率建模并估计每一个个案的似然性。在第二阶段，为不能观测的、假定为正态分布的潜在应答变量建立一个共同的测量尺度。应答的阈值被假定不随时间变化而变化。在第三阶段，一个增长模型被拟合到具有尺度的潜在应答变量上。目前，仅有几个软件能够估计此类模型，如 Mplus 与 Mx。如果有缺失数值，Mx 是唯一的选择。Jöreskog(2002) 用 LISREL 描述了一个类似的程序，涉及从拟合模型中得到一个多分格协方差矩阵以及从原始的定序数据中估计得到均值。

在方法论文献中，对于分类变量进行潜变量增长曲线建模的应用正受到越来越多的关注。例如，Liu 与 Powers(2007) 最近描述了一种在潜变量增长曲线

模型框架中对零膨胀计数数据建模的方法。在可预期的将来，会有更多有意思的进展。要了解更多的对二分变量和定序变量建立模型的方法，参见 Jöreskog (2002)、Mehta 等人(2004)以及 B. Muthén 与 Asparouhov(2002)的著作。

## 在变化中对因果关系建模

讨论至此，截距和斜率的(协)方差都是“无结构的”，即允许它们自由共变。我们也可以相应地选择估计变化方面的直接影响，也就是说，我们可以对变化建模，把它看做另外一些变化的函数。例如，时间尺度集中在测量的最后一次，如果我们能对“结束点”建模，把它作为变化的函数，我们就可以对截距和斜率进行回归。另外，我们也可以在起始点集中，把斜率作为截距的线性函数来建模。当然，我们要特别谨慎，原因应该在逻辑上早于结果，否则用截距对斜率回归就会在原因和结果上不匹配，除非原始时点发生在第一次测量以前。

在一些情况下，研究者会将参与者随机地安排到至少两个组中，并对一些他们感兴趣的结果进行多次重复测量，例如干预性研究。Muthén 与 Curran (1997)创造性地利用了 SEM 的这个特征对实验效应进行建模。他们建议在两个小组中拟合同样的曲线，限定各组中的线性增长参数相等(见图 3.3)。第二个斜率因子(Tx 斜率)为实验组所特有而被加入到模型中。控制组提供了一个基线轨迹，因此实验组的轨迹能够与之比较。任何和第一个斜率相关的、被观测到的额外变化都可以被看成实验的影响。模型的一个重要的方面是额外的实验斜率因子可以对截距因子(初始状态)进行回归分析，并允许检测截距和实验的交互影响。例如，在干预性研究预防外部行为的时候，如果母亲—子女亲密度的值在起始状态较低，可能会在之后的干预中得到更大的益处。在 Muthén 与 Curran 的模型中，截距因子会影响第三个潜变量(干预)，而后者又会影响随时间变化的测量变量，但此影响仅仅发生在实验组中。

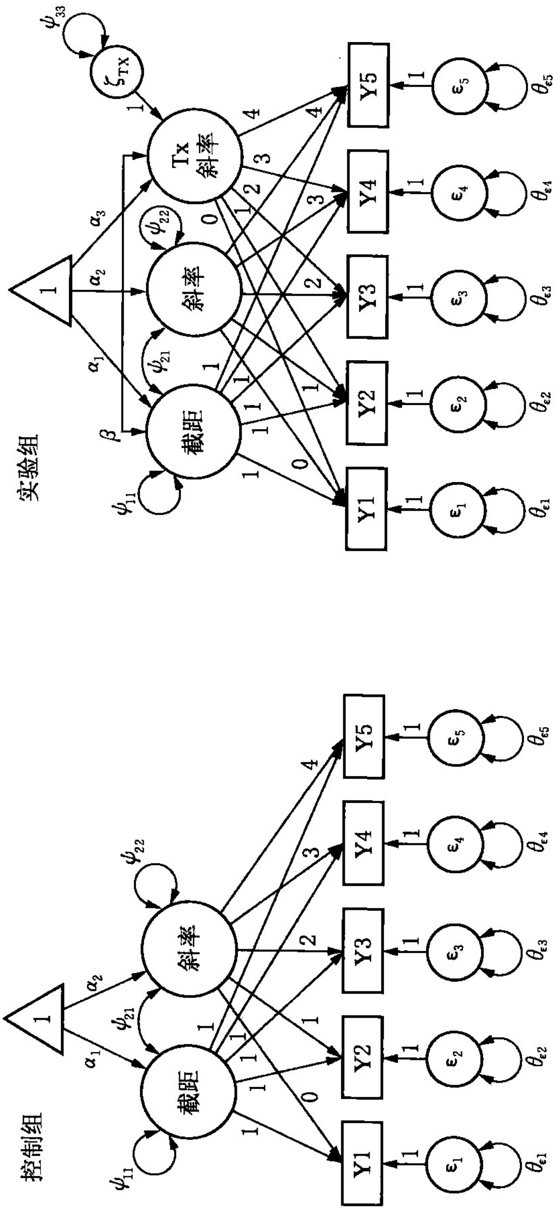



图 3.3 Muthén 与 Curran(1997)干预性研究的双组增长模型

## 总结

在这一章中，我们描述了几个潜变量增长模型的特别拓展，说明潜变量增长模型作为一个分析工具是具有一般性和普遍性的。

根据研究者的要求，结合这里讨论的模型的特征，用其他方法扩展这些模型也是可能的。例如，将增长混合模型应用到队列序列数据中，并检测 TVCs 的影响，或者将 TVCs 看成并行过程 ALT 模型外部预测变量也是可行的（Bollen & Curran，2004；Curran & Willoughby，2003）。Simons-Morton、Chen、Abroms 和 Haynie(2004)对 3 个并行过程（青少年吸烟、朋友吸烟和父母介入）建模，包括非时变的预测变量并设定截距对斜率在过程之内和过程之间的效应。Sayer 和 Willett(1998)结合分阶段和并行过程增长模型，对两个组同时进行拟合。McArdle(1989)描述了一个多组并行过程模型。最后，虽然我们没有对以上这些模型进行说明，但是把截距和斜率看成结果变量的预测变量是十分简单的。例如，检验某种技能是否能预测个人在多年后能力的差异这一假设可能是重要的。Muthén 与 Curran(1997)描述了一些模型，其中截距和斜率因子被用来预测一个远端结果或者与截距和斜率因子相关的重复测量变量。

潜变量增长曲线模型框架允许设置和检验几个不同种类的交互和调解的假设。例如，研究者能够对两个或者多个外生自变量的斜率（Curran et al.，2004；Li Duncan & Acock，2000；Preacheret et al.，2006）在初始状态和决定结果的时段（Muthén & Curran，1997），或者在时段和 TVC 之间（见第 2 章模型 9 的讨论）的交互进行假设检验。通过应用多层结构方程模型，潜变量增长模型也可以被扩展到三层次（或者多层次）的多层数据分析中（Mehta & Neale，2005；B. Muthén，1997）。另一方面，作为潜变量增长模型中的特殊例子，拟合简单的 ANOVA、MANOVA 和单一模型也是可行的（Meredith & Tisak，1990）。

# 第 4 章 | 潜变量增长模型和多层模型的关系

纵向数据建模的潜变量增长曲线模型的一个有效的替代方法是多层模型(MLM; Bryk & Raudenbush, 1987, 1992; Goldstein, 1995; Hox, 2002; Kreft & de Leeuw, 1998; Raudenbush & Bryk, 2002)。多层模型,也叫做"分层线性模型"(HLMs; Bryk & Raudenbush, 1992; Raudenbush & Bryk, 2002)、"混合模型"、"方差分子模型"或者"随机系数模型"。这一模型的一个重要的能力是在分析嵌套数据时(例如,当学生被嵌套在学校之中时),它允许多级方差的区分。对 MLM 模型的一些介绍可以在 Hofmann(1997)、Kreft 与 de Leeuw(1998)以及 Luck(2004)的著作中找到。MLM 和 LGM 分别起源于不同的统计研究和理论,它们各自有专门的术语和研究问题的标准体系。但是,现在越来越明显的是,这两种模型在很多情况下存在很多重复之处。在本章,我们主要讨论它们的重叠部分。

在多级嵌套的数据中,最低的一层总是被叫做"第一层",更高的层有更高的次序。在很多情况下都可以看到嵌套数据,例如,学生(第一层)被嵌套在学校(第二层)之中,工人被嵌套在产业之中,病人被嵌套在治疗小组之中。结果变量通常是在第一层测量的,但是他们的预测变量可以存在于第一、第二甚至更高层。一般来说,每个组(第二层的单位)会在解释变量和结果变量上有不同的关系。解释变量在各组中的系统变化通常被用来解释这种组内的不同关系。例如,私立学校的学生可能比公立学校的学生有更好的考试成绩,因此研究者也许想检验不同的干预方式对消除这个组间区别的功效。在这个例子中,MLM 可以被用来研究不同的矫正阅读计划在不同的环境中可能会有的不同影响。

## MLM 在重复测量数据中的应用

使用多层模型研究组内的嵌套和研究增长曲线之间有一个很简明的概念关联(Diggle, Liang & Zeger, 1994; Longford, 1993)。在重复测量的设计中，测量的时点(第一层)嵌套在个人之中(第二层)，一旦测量时点被作为第一层的测量嵌套第二层之中，多层模型分析对纵向数据的处理就同其他多层模型的应用一样了。

一般来说，研究者通常以估计一个随机效应的方差模型——一个没有解释变量的模型——来决定多层模型是否比传统的最小二乘法更好。这个模型把结果变量的方差分成因时间而产生的差异(个人在不同重复测量时点的差异，即第一层)和不同个人之间的差异(个人之间的差异，即第二层)。如果有一大部分人之间的差异需要被解释，研究通常会逐步把预测变量加到公式中。这种做法允许在第一层变量时间上定义结果变量的变化，并且可以通过加入在个人之间变化的第二层的解释变量来进一步建模。

这些解释变量或有固定系数，或有随机系数。固定系数是在估计上忽略与第二层变化相关的系数。例如，一个研究者想研究随着时间变化，儿童每周零用钱(第二层)对儿童的数学成绩(第一层)的影响，设定零用钱收入作为一个固定系数，将产生一个零用钱对儿童学习成绩的总体平均影响。但是，设定零用钱的斜率系数为随机系数，将会额外地产生一个零用钱对儿童数学成绩影响上的差异的估计，这个估计又可以用别的变量来建模。在 MLM 的几个优点中，随机系数估计的方差可以通过更高层次的解释变量来预测，这是一个很大的优势。一般来说，在 MLM 中，这些解释变量先被作为固定系数包括进去，如果模型的拟合度能被大大提高，就可以使系数随机变化。

使用多层模型检验多重测量数据的一个明显优势是，假定缺失数据是随机的，即使在一些场合下有部分缺失数据的个人，它仍能够使用所有可获得的数据。此外，用 MLM 模型不需要假定每一个研究对象都在统一时点或者年龄上测量，甚至测量的次数也可以是不一样的。鉴于这些优势，如果研究者在

多重测量数据研究中想要得到一个可靠的结果，MLM 似乎是一个很好的选择。

## 模型设置

一般来说，用来分析线性变化的 MLM 模型的结果变量 $y$ 会作为时间的线性函数[22]：

$$y_{ij} = \beta_{0j} + \beta_{1j}(time_{ij}) + \varepsilon_{ij} \qquad [4.1]$$

$$\varepsilon_{ij} \sim N\{0, \sigma_\varepsilon^2\} \qquad [4.2]$$

在此，$\beta_{0j}$是个体 $j$ 的截距，$\beta_{1j}$是个体 $j$ 的斜率，$\varepsilon_{ij}$是个体 $j$ 在时点 $i$ 的残余项（服从均值为 0、方差为 $\sigma_\varepsilon^2$ 的正态分布）。截距和斜率一般都被看成随机变量，关于个体之间的方差建模如下：

$$\beta_{0j} = \gamma_{00} + u_{0j} \qquad [4.3]$$

$$\beta_{1j} = \gamma_{10} + u_{1j} \qquad [4.4]$$

$$\begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix} \sim N\left\{\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \tau_{00} & \\ \tau_{10} & \tau_{11} \end{bmatrix}\right\} \qquad [4.5]$$

在设定这两个参数为随机数据时，截距（$\beta_{0j}$）和斜率（$\beta_{1j}$）项被设定为一个包含固定效应的函数，即样本中所有人特征的平均效应加上每个人在固定效应基础上的偏离。在方程 4.3 和方程 4.4 中，$u$ 代表这些偏离（$u_{0j}$为偏离平均截距的部分，$u_{1j}$是偏离平均斜率的部分）。这些偏离项的方差和协方差（$\tau_{00}$、$\tau_{11}$和 $\tau_{10}$）通常是有理论旨趣的部分，例如，研究者可能想研究那些起初数学能力比较高的儿童的成绩是不是比那些起初比较差的儿童的成绩提高得更快。截距与斜率之间正的协方差显示那些起初成绩好的儿童随着时间的推移，其提高也更快。把方程 4.1、方程 4.3、方程 4.4 结合，会产生一个扩展方程：

$$y_{ij} = \gamma_{00} + \gamma_{10}(time_{ij}) + u_{0j} + u_{1j}(time_{ij}) + \varepsilon_{ij} \quad [4.6]$$

方程 4.6 显示了一个多层的回归方程可表示为固定成分——$\gamma_{00} + \gamma_{10}(time_{ij})$ 和随机成分——$\mu_{0j} + \mu_{1j}(time_{ij})$ 的总和。这个模型估计了两个固定效应($\gamma_{00}$和 $\gamma_{10}$)和 4 个随机效应——个人在固定的截距和斜率的偏离的方差和协方差($\tau_{00}$、$\tau_{11}$ 和 $\tau_{10}$)以及模型中任何剩下的、不能解释的方差($\sigma_{\varepsilon}^2$)。基于对这些截距和斜率的方差和协方差的估计，研究者可以对研究问题进行合适的推论。

与传统的回归类似，MLM 模型一方面能够分别估计个人层次的影响和同一个人在不同测量上的影响，另一方面，它还能够研究变化之间的关系，因而作用更强大。当将多层模型应用到多元变化中时，或当我们对起始的水平(截距)和变化速度(斜率)间的关系感兴趣时，后一个特点尤其有用。当然，基于这个标准的随机截距、随机系数的模型针对特别例子的许多扩展也是可能的。例如，研究时变变量和非时变变量是可能的，也可以添加和检验一些模型参数的限定，并估计多元模型。

## 参数估计

ML 和 FIML 估计通常被用来在多层模型中获得参数估计。正如在结构方程模型中一样，ML 估计在多元正态分布的假设前提下会产生一套参数估计，以使可观测到的数据成为最可能的结果。不同的统计软件在获取最大似然估计时会采取不同的方法。例如，MLwiN(Rasbash et al.，1999)用一个广义迭代最小二乘法获取所有参数的最大似然估计值。另外，当最大似然估计被应用到最小二乘化的残差中时，这种方法被称为“有限制的最大似然估计”(REML)。在 MLwiN 中获得 REML 解的算法是有限制的广义迭代最小二乘法。较之 MLwiN，HLM(Bryk，Raudenbush & Congdon，1996)使用将期望值最大化的算法提供一个经验性的贝叶斯估计，当其满足正态分布假定时，它相当于广义最小二乘法的估计。FIML 在似然估计中既包括固定参数，也包括随机参数；而

REML 则是估计模型中固定效应消除后的方差成分(Hox，2000、2002)。Hox 进一步指出，因为这一点，当方差差异检验涉及不限制固定参数的情况时，最大似然估计不应当被用来进行方差差异检验(也可参考 Singer & Willett，2003：90、118)。

## 模型评估

同最小二乘回归一样，多层模型的假设通常会通过利用固定统计上或实际中回归系数的显著性来评价，虽然随机系数的方差和协方差经常是我们感兴趣之处。参数显著性的一个非正式检验是将其点估计值除以它的标准误，如果这个比率超过 2.00(在无限大的样本中是 1.96)，估计所得参数就可称为“在 0.05 显著水平上与 0 显著不同”。关于多层模型的显著性检验的更详尽的讨论，请参见 Hox(1998)的有关论述。

多层模型的总体拟合度的信息也是可获得的，它会被作为一个“偏离统计量”报告出来，定义为 $-2\ln L$。在这里，$L$ 是似然函数在收敛处的似然值(Hox，2002；Kreft & de Leeuw，1998；Li，Duncan，Harmer，Acock & Stoolmiller，1998)。偏离统计量并不能为模型总体的拟合度提供一个容易解释的度量，因为这个指标和样本数目的大小密切相关。然而另一方面，在嵌套的模型中，偏离统计量之间的差异服从卡方分布，自由度等于嵌套模型之间参数数量的差异(Goldstein，1995；Hox，2000、2002；Li et al.，1998)。

## MLM 和 LGM 重合的方面

MLM 和 LGM 可以用于涉及重复测量数据的很多相同的情况。事实上，两者因为在方程式的层次上具有相似的模型表达，所以在很多情形下会产生同样的结果。因此，认为 MLM 和 LGM 相似也不是什么新的提议。这两种方法重叠的各个方面在各种文献中已经被多次说明了。在两个框架中，相应的重复测量模型可以被设定来获得同样的参数估计(在可容忍的误差内)(Bauer，2003；

Bauer & Curran, 2002; Chou, Bentler, & Pentz, 1998; Curran, 2003; Hox, 2000、2002; Khoo & Muthén, 2000; Li et al. , 1998; MacCallum & Kim, 2000; MacCallum et al. , 1997; Raudenbush, 2001; Rovine & Molenaar, 2000; Stoel, 2003; Wendorf, 2002; Willett & Sayer, 1994)。例如,在已经描述过的基本的线性增长模型中,多层模型中所有我们感兴趣的 6 个参数在潜变量增长模型里都可以找到相应的参数。具体说来,平均截距估计在两个模型中分别由 $\gamma_{00}$ 和 $\alpha_1$ 表示(见表 4.1)。同样,随时间变化而产生的均值变化,即斜率参数,在这两个模型中分别由 $\gamma_{10}$ 和 $\alpha_2$ 表示。截距和斜率的方差、协方差以及方差的分布也都能在两种方法中得到表达。虽然在结构方程模型和多层模型中的估计方法是不一样的,但它们都会使用所有可以利用的信息产生参数的最大似然估计值(包括它们的标准误)。因此,对于在两种框架中都有相同代表的模型,我们可以合理地期待多层分析中的参数估计值和潜变量增长模型中的参数估计值是一样的。[23]表 4.1 包含一个实例,用母亲—子女亲密度数据和模型 4 来说明这一问题。其中多层分析是使用 MLwiL1.1 完成的,参数估计值和标准误与表 2.6 所显示的一样。

MLM 和 LGM 的另一个有趣的相似性也值得注意。在 MLM 的传统中,通常会得到随机系数的经验贝叶斯估计值,即基于每一个第二层单位的模型得到的截距和斜率的估计值,这通常是为了绘图和诊断的目的。这些参数也被称为"后验均值"(因为它们是给定数据的随机系数的后验分布的均值估计)或"收缩估计值"(因为它们会与第二层单位的可靠性成比例地朝着截距和斜率的均值的方向收缩)。在潜变量的传统中,经验贝叶斯估计值相当于因子分数,即潜变量的后验估计值。在 LGM 中,这些估计值是模型暗含的截距与斜率的值,根据每个个体(或第二层单位)贡献的可靠性加权得到。我们认为,在 LGM 中计算因子分数是没有实际价值的,主要是因为它们是有偏估计值,不能准确反映个体的轨迹。而且,Bartholomew(2007)指出,研究者不可能确定地将因子分数分配给个体,因为对于每个个体,模型都暗含了一个这一分数的分布,而因子分数则表示这些分布的均值。

表 4.1 与模型 4 对应的多层模型的参数估计

| 多层模型 | 潜变量增长模型 | 解　释 | 估　计 |
| --- | --- | --- | --- |
| $\hat{\gamma}_{00}$ | $\hat{\alpha}_1$ | 平均截距 | 38.00(0.08) |
| $\hat{\gamma}_{10}$ | $\hat{\alpha}_2$ | 平均斜率 | −0.36(0.02) |
| $\hat{\tau}_{00}$ | $\hat{\psi}_{11}$ | 截距方差 | 2.98(0.29) |
| $\hat{\tau}_{11}$ | $\hat{\psi}_{22}$ | 斜率方差 | 0.14(0.02) |
| $\hat{\tau}_{10}$ | $\hat{\psi}_{21}$ | 截距—斜率协方差 | 0.25(0.06) |
| $\hat{\sigma}_{\varepsilon}^{2}$ | $\hat{\theta}_{\varepsilon}$ | 扰动项方差 | 3.70(0.10) |

## MLM 和 LGM 的不同之处

对于这里考虑的纵向数据，LGM 和 MLM 在数学上是等同的。两种方法中一些可以觉察到的差别是由计算它们的统计软件(Raudenbush & Bryk，2002)和历史传统上的差别造成的。虽然每种方法比起另外一种的确都有一些灵活性方面的局限性，但是以前的一些局限性已经随着软件水平的提高而消失。上一节的目的是说明两种模型应用于分析纵向数据时重合的地方，但在有些情况下，一些应用可能在一个框架下可行，在另外一个框架下就不可行。这些方法上不能重叠的部分是由目前一些操作软件的执行局限性所致的。例如，比起 MLM，LGM 的一个显然的优势在于，LGM 有能力设定一个测量模型，也就是说，重复测量不需要像在 MLM 模型中那样必须避免测量误差。在 LGM 中，每一个重复测量的变量可以被设定成一个有很多指标的潜变量(Curran，2000；Khoo & Muthén，2000)。然而，假定重复测量的潜变量因子结构不随时间变化的模型较之测量模型有更大的统计效力，但理论上来说，测量模型也能够被纳入 MLM 软件包中，如 Mplus 就能够估计多层次结构方程模型(L. K. Muthén & Muthén，1998—2006)。

MLM 模型的第二个局限性是它不能够在设定上让模型参数去预测别的变量。而 LGM 不仅仅把这些参数变化作为参数，更可将其作为变量，所以 LGM 没有这些缺陷(Bauer & Curran，2002；Hox，2000)。目前许多不能够通过

MLM 实现的复杂的模型却可以在 LGM 模型中轻易实现。一个最近文献中的例子就是在第 3 章讨论过的、对随时间变化的群体间差异进行假设检验的应用（Muthén & Curran，1997）。目前，纯粹的 MLM 还不能估计像 Muthén 和 Curran 那样的通过建构一个潜变量来影响另一个潜变量建构的模型。

MLM 模型的第三个局限在于它不具有 LGM 模型能够估计因子负载的能力，这个能力在未设定轨迹的增长模型（第 2 章的模型 11）和结构潜变量曲线模型 SLC 中非常有用。在 MLM 框架中，时间被当做一个数值已知的变量（MacCallum et al.，1997）。

虽然 LGM 模型有一些独特的优点，但这并不意味着 MLM 模型没有它独到的长处。MLM 模型较之 LGM 模型一个明显的长处是，MLM 模型能够处理两个以上层次的模型。例如，一个数据包含学生这一层的多重测度，学生又嵌套在班级之中。这样的模型在 MLM 中十分简单易行，但是在 LGM 模型中就很难实现（Curran，2003；Muthén & Muthén，1998—2006）。关于多层结构方程模型的一个较好的概述，参见 Hox（2000）、Metha 与 Neale（2005）的著作。另外，虽然 SLC 模型能够近似地对时间的复杂的非线性函数建模，但这样的函数在 MLM 模型框架中不需要近似就可以直接建模。这取决于使用什么样的软件包。

在这节中，我们已经试图列出存在于 MLM 和 LGM 中有关模型设定的一些差异。如果我们把讨论的内容展开一些，还可以探讨很多别的问题，包括由历史根源导致的差异。例如对于我们已经描述过的模型，一方面存在是否使用相应的拟合模型的统计指标的差异，另一方面，修正模型指标常常在 LGM 与 SEM 中被视为投机（MacCallum，Roznowski & Necowitz，1992），而在 MLM 中，解放参数则被认为是可以接受甚至是受到鼓励的。除了注意到在这些不同传统的分支中，历史原因在很大程度上决定了研究者认为什么做法是可以接受的，我们也要注意判断何种模型的优越性在很大程度上依赖于软件发展的分析传统，而不一定是技术上的优越性。

虽然 MLM 和 LGM 模型分别起源于不同的理论与实践传统，但这两种模型对于研究者来说都是有用的工具，它们都能为关于随时间变化的问题提供相

似的答案。多亏最近的软件发展和估计技术的进步，以前被认为在一种方法或另一种方法中不可能实现的方面，现在也变为可能。例如，缺失数据估计已经使 LGM 这种方法很接近 MLM 模型的应用。另外，MLM 软件现在有能力对不同的误差项结构建模，这在以前是 LGM 的一个优势。MLM 和 LGM 之间的差异正在迅速缩小，但是，这两种方法现在依然是不同的，研究者需要慎重地考虑使用哪种方法来进行模型检验和发展。随着模型设定框架和软件的发展，预计这两种模型的区别最终将彻底消失。

## 软件

目前，大量的统计软件包都可以做 MLM。专用的软件包括 aML、HLM 和 MLwiN。几个一般的统计软件包括 R、SAS、SPSS、Stata、SYSTAT 和 SEM 程序。EQS、LISREL 和 Mplus 现在也有能力估计多层模型。在这些软件中，R、SAS、Splus 和 Stata 包含能够估计非线性多层模型的程序。对于二分变量和定序变量作为结果变量的模型，则需要 REML 和 FIML 之外的方法(例如，最大边际似然估计、惩罚性准似然估计、经验性贝叶斯估计或者马尔科夫链—蒙特卡罗估计)，因此对于这些类型的数据，经常需要特别的软件。aML 可以在网上免费获取，此外它还能够处理几种不标准的数据类型和模型，如包含截断数据和删失数据的回归模型、多层风险模型、定序或者计数数据、曲线模型和很多其他模型。MIX(一个独立的程序集，或者 SPSS 和 SAS 的宏命令)也能够估计很多不同数据类型的多层分析模型。如果研究者想要对定序变量、计数变量或者一般的非线性数据建模，我们建议研究这些 LGM 之外的选择。更多关于多层模型的软件信息可以参考 de Leeuw 与 Kreft(2001)以及 Snijders 与 Bosker(1999)的著作，还有上面提到的统计软件包的使用说明。

# 第 5 章 | 结论

在本书中，我们对潜变量增长曲线模型做了一个介绍：用数据说明各种应用，描述几种模型的拓展和更高级的模型在 LGM 框架中的应用。我们说明了如何对比简单的线性模型，并对更高级的模型进行建模、设定和评估，包括用群体作为预测变量、设定相关增长曲线以及包含队列次序设计。我们展示的各种可能的模型表明，这类技术既是灵活的，也是强有力的工具。LGM 作为一个应用统计的领域，在最近的方法理论文献中正引起越来越多的注意，从各个方面完善这个方法的文献也正在迅速累积。这本书为那些正在寻求适应这个领域迅速发展的研究者提供了一个概念、理论和应用上的基础。

LGM 框架的一个最令人注目的特点是它具有把现存的模型用一些新的方法进行扩展的潜力。我们尤其肯定的是，结构潜变量曲线方法在应用实践中有很大的潜力。在实践中应用复杂的曲线模型的唯一障碍是缺乏容易获得和使用的软件，这个问题一定会在将来得以解决。现在，大量的相关领域的研究也在迅速发展，包括交互潜变量的增长曲线模型(Li et al.，2000)、绘图和探索条件增长曲线的简单效应(Curran et al.，2004；Preacher et al.，2006)以及对增长混合模型的进一步发展和完善(Bauer & Curran，2003a、2003b、2004)。除此之外，很多最近的研究也在探究对缺失数据的处理以及对分类数据的研究。在一些其他领域，研究者也有机会对 LGM 进行拓展。例如，虽然 $\varepsilon$ 中的干扰项通常被设定成相互独立的，但是这种限制也可以是不必要的(Browne，1993；McArdle，1988)。允许干扰项服从一个一阶的自相关结构是很平常的，通常相邻干扰项的协方差被设置成自由参数(Browne，1993；Meredith & Tisak，1990；Willett & Sayer，1994)。另外，设定一个基于误差协变量的因果结构也

是可能的。

虽然我们关注的问题是在频率统计框架中的LGM模型，但计算机技术的提高已经使得贝叶斯方法比过去更有用。贝叶斯方法在LGM的应用包含对定序变量和二分变量的建模、非线性增长模型的建模、多样本和混合模型建模等，还有对于可忽略的或不可忽略的缺失数据进行拟合、对非正态数据的拟合以及以上这些在标准结构方程模型中往往会引发问题的模型的组合。贝叶斯统计的一个重要标志，即把先验的信息包括在模型中，是可能的。简言之，贝叶斯统计方法在LGM中的应用是灵活强大并具有潜力的，但是它往往会要求使用者具备更高的统计知识。关于这一新兴方法的更多信息，有兴趣的读者可以参考Lee(2007)，Scheines、Hoijtink与Boomsma(1999)以及Zhang、Hamagami、Wang、Nesselroade与Grimm(2007)的著作。

虽然这本书关注的是潜变量增长曲线模型，但是这种模型不是唯一的探索纵向数据的潜变量方法。和LGM相关的技术包括探索性纵向因子分析(Tisak & Meredith, 1989、1990)和使用纵向结构方程模型探索随时间变化的测量信度、效度与稳定性等(Tisak & Tisak, 1996、2000)。自回归模型和自回归交叉滞后模型在很多情况下是对增长曲线模型的可行替代(参见Bollen & Curran, 2004; Curran et al., 1997)。上一章讨论的多层模型在分析纵向数据中也有广泛的应用。

虽然我们试图提供尽可能必要的信息以使读者对于LGM模型有一个完善的基本了解，但是我们的目标还远未实现。我们的目标是为基本的LGM方法提供一个概述，展示它的应用，并描述方法上的一些最新进展。在这些主要的问题之外，一个更广泛的建议是，通过规划明确地包含时间的模型可以获得许多关于数据本质的真知灼见(Collins, 2006)。正如Anderson(1993:929)所说："一个真正的纵向研究是无可替代的。"LGM分析代表了这种研究的一个类别，它能够帮助研究者形成假设，探索变化形态和相关的趋势。

我们特意将参考书目列得很长，因为这本书的另一个目的是为有兴趣的研究者提供一个入门参考。关于在这本书中讨论的特定问题的更多信息，我们推荐读者去阅读相关的引用文献。对于LGM模型的介绍，请参考Bryne和

Crombie(2003)、Chan(1998)、Curran(2000)、Curran 和 Hussong(2003)、Hancock 和 Lawrence(2006)以及 Willett 和 Sayer(1994)的著作。Singer 和 Willett(2003:第 8 章)对 SEM 和 LGM 提供了很好的介绍。关于 SEM 更复杂和详细的论述,我们强烈推荐 Bollen(1989)的著作;关于 LGM,我们推荐 Bollen 和 Curran(2006)的著作。

# 附 录

想更多地了解潜变量增长曲线模型的研究者还可以从网上获得资源，下面我们就将简要列出这些资源。

## 绘图和探索潜变量增长曲线模型中的交互项

在普通线性回归中，交互(调节)效应通常是通过描绘交互图来实现。为了便于理解，交互图展现出在调节变量 $M$ 的特定值下 $Y$ 对 $X$ 的回归结果。LCA 框架的一个主要优势在于，表示截距和斜率的因子可以是其他模型中的内生(因)变量，如模型 6 那样。当这些变化可以通过外在变量 $M$ 来预测时，重复测定变量 $Y$ 受到时间的影响就随着 $M$ 而变化，这一点非常类似普通回归中 $Y$ 对 $X$ 的回归受到 $M$ 的影响。反之，我们同样可以说，$Y$ 受到 $M$ 的影响随着时间而变化。这种效应被称为“交叉水平交互作用”。

回归方法中，通常用于描述交互作用的方法(Aiken & West, 1991)也可以用于 LGM 框架中(Curran et al., 2004; Preacher et al., 2006)。我们不仅可以画出任意 $M$ 值下，模型中 $Y$ 随着时间变化的平均轨迹，甚至还可以描绘出任意情况下 $Y$ 对 $M$ 的回归图。绘制交叉水平交互的工具可以参考 http://www.quantpsy.org，其中还包括指南和完整的实例。

## 描绘个体轨迹

LGM 分析中的大部分信息都是总结性的，输出结果通常的形式包括平均

趋势、这些均值的变化和某些不符合这些趋势的个案。Carrig、Wirth 和 Curran (2004)曾用 SAS 中的宏(macro, OLS*traj*)来绘制群体层面和个体层面的轨迹图。OLS*traj* 可以作为起始步骤辅助进一步的 LGM 分析,也可以作为描述数据的方法。除了提供个体轨迹图之外,OLS*traj* 还可以提供估计参数的群体层面的直方图和箱形图,同时还可以输出个体参数估计值。OLS*traj* 和其手册可以在 http://www. unc. edu/~curran/olstraj. htm 上得到。

## 程序语句

本书中所有分析的 LISREL、Mplus 和 Mx 语句,都可以在第一作者的网站(http://www. quantpsy. org)上找到。

# 注释

[1] http://www.quantpsy.org/.

[2] 对于熟悉在结构方程模型中 LISREL 框架下模型设定的读者，这里的数学表达式使用的是"全一 $y$"模型，即所有的潜变量都被认为是内生的。

[3] 矢量由下划线表示，而矩阵由黑体字表示。

[4] 重要的是要记住 $\boldsymbol{\Theta}_\varepsilon$ 中的元素并不表示通常意义上的误差方差。更确切地说，$\boldsymbol{\Theta}_\varepsilon$ 中的第 $t$ 个方差反映了一个线性模型在多大程度上没能在 $t$ 个时点获得个体的分数。这种变化性大多是误差造成的，但也可能是由于其他变量预测的影响。

[5] 另一方面，Stoolmiller(1995)建议在多项式中将时间变量对均值对中，以避免由多项式各项的线性依赖而造成的估计问题。

[6] 当"缺失"并不依赖于 $x$ 或其他变量的可观测或不可观测的反应时，缺失数据(对于 $x$)是完全随机缺失。当"缺失"在控制数据中的其他变量的条件下而并不依赖于 $x$ 的可观测或不可观测的反应时，$x$ 的缺失数据是随机缺失(Allison，2002)。

[7] 据我们所知，Mplus 是唯一在有数据缺失时仍然计算拟合指数的结构方程模型软件。

[8] 可能是由于增长曲线模型的高度限定性，根据总体拟合标准，它们的拟合常常较差。但是，有时在个体增长曲线模拟得较好的情况下，模型也会拟合较差。Coffman 和 Millsap(2006)建议用个体拟合标准来补充总体拟合指数。

[9] 这一自由度的公式假设样本数据由均值与协方差构成。增长曲线模型几乎总是拟合到均值与协方差。如果没有对均值结构建模，那么自由度为 $[p(p+1)/2]-q^{*}$。

[10] 我们为了教学的简明易懂而使用完全数据。而且，使用完全数据可以获得完整的结构方程模型拟合指数。http://www.quantpsy.org/提供了接下来使用的分析数据以及计算机语句。

[11] LISREL 可以从科学软件国际的网站上(http://www.ssicentral.com/)获得；Mx 可以从弗吉尼亚联邦大学的网站上(http://www.vcu.edu/mx/)获得；Mplus 可以从 L. K. Muthén 与 Muthén 的网站上(http://www.statmodel.com/)获得。

[12] 在整本书中，我们为了教学的简明易懂而把扰动项方差限定为相同。而且，设定母亲一子女亲密度的残差变异性一直随时间保持稳定是合理的。如果有足够的重复测量来识别这些参数，是不需要这一相同的限定的。实际上，LGM 方法的一个强项就是我们可以通过估计不同时点的残差方差来明确地对异方差性或其他各种各样的扰动项协方差结构进行建模(Willett & Sayer，1994)。但是，在其他条件相同的情况下，我们建议应尽可能地先对同方差性建模以获得最大化的简约性。而且，由于允许残差方差在不同时点不同，有时会掩盖或"吸收"轨迹中的非线性，产生欺骗性的好拟合。

[13] 我们的网站(http://www.quantpsy.org/)提供了这一模型以及接下来的大多数模型的程序语句。

[14] 这就是我们前面提到的嵌套模型检验。在这一检验中，两个模型的卡方值的差异本身被当做一个卡方统计值，它的自由度等于两个嵌套模型的自由度的差异。

[15] 需要说明的是，在这一过程中并没有真正估计单独的截距与斜率，相反，这一模型将一个多元正态分布加于潜变量并得出这些分布的均值、方差和协方差的估计值。

[16] 孔径或许可以在某些结构方程模型软件中通过把 $\psi_{21}$ 限定为 0，并把斜率因子限定为它们的原始固定值减去一个孔径参数而作为模型参数的直接估计。

[17] 在这个模型以及后面的模型 9 中，我们会使用 $\beta$ 参数，它们是第 1 章中没有讨论的完全结构方程模型的路径系数矩阵(**B**)的元素(Bollen，1989)。

[18] 因为这一相似性，多层模型中的跨层交互可以按照 Aiken 和 West(1991)的指导被分解、检验和绘图。更多的讨论参见 Curran、Bauer 和 Willoughby(2004)以及 Preacher、Curran 和 Bauer(2006)的著作。

[19] 这两种检验队列效应的方法同模型 5 和模型 6 中所讨论的两种检验变化的外生自变量的效应是可直接比拟的。

[20] 以似然值为基础的置信区间是由确定为了使模型拟合度下降某个给定的数量时，一个参数必须取的值来计算的。例如，95%的置信区间是由把参数值一点一点移离最大似然估计值(每一次都重新对模型进行最优化)，直到最大似然估计函数增加 3.84 个卡方单位(3.84 是自由度为 1 时卡方的临界值)。由于需要大量的中央处理器密集地重新最优化，这一设定置信区间的方法会非常费时，但是以似然值为基础的置信区间较之标准误有好几个优势，后者以正态分布(对方差参数往往不正确)为前提，需要随重新参数化而变化的 $t$ 检验，并且有时会对边界参数得出荒谬的结果。虽然标准误在报告参数估计值时仍然是必要的，但我们预测因为它们可取的特性，以似然值为基础的置信区间将变得更受欢迎，特别是随着计算机价格的下降和计算效率的提高。

[21] 为了能识别这一模型，两个负荷必须被限定。任意两个负荷都可以，只要它们被限定为不同的值。

[22] 我们的标记方法与 Kreft 和 de Leeuw(1998)的相似，用下划线表示随机系数。在之前讨论的潜变量增长曲线的矩阵中，下划线表示矢量。

[23] Hox(2000)为这一紧密的相似性提供了一个证明，他发现在数据完全随机缺失或随机缺失的情况下，对于所有的方法，前者的参数估计值更为接近。

# 参考文献

Aber, M. S. & McArdle, J. J. (1991). "Latent growth curve approaches to modeling the development of competence." In M. Chandler & M. Chapman(eds.), *Criteria for competence: Controversies in the conceptualization and assessment of children's abilities*. Hillsdale, NJ: Lawrence Erlbaum.

Aiken, L. S. & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions*. Thousand Oaks, CA: Sage.

Allison, P. D. (1987). "Estimation of linear models with incomplete data." In C. C. Clogg(ed.), *Sociological methodology*. San Francisco: Jossey-Bass.

Allison, P. D. (2002). *Missing data*. Newbury Park, CA: Sage.

Anderson, E. R. (1993). "Analyzing change in short-term longitudinal research using cohort-sequential designs." *Journal of Consulting and Clinical Psychology 61*:929—940.

Arbuckle, J. L. (1996). "Full information estimation in the presence of incomplete data." In G. A. Marcoulides & R. E. Schumacker(eds.), *Advanced structural equation modeling: Issues and techniques*. Mahwah, NJ: Lawrence Erlbaum.

Arbuckle, J. & Friendly, M. L. (1977). "On rotating to smooth functions." *Psychometrika 42*: 127—140.

Arbuckle, J. L. & Wothke, W. (1999). *AMOS 4.0 user's guide*. Chicago: SPSS, Inc.

Arminger, G., Wittenberg, J. & Schepers, A. (1996). *MECOSA 3 user guide*. Friedrichsdorf/Ts, Germany: ADDITIVE GmbH.

Baer, J. & Schmitz, M. F. (2000). "Latent growth curve modeling with a cohort sequential design." *Social Work Research 24*:243—247.

Baker, G. A. (1954). "Factor analysis of relative growth." *Growth 18*:137—143.

Bartholomew, D. J. (2007). "Three faces of factor analysis." In R. Cudeck & R. C. MacCallum (eds.), *Factor analysis at 100: Historical developments and future directions*. Mahwah, NJ: Lawrence Erlbaum.

Bauer, D. J. (2003). "Estimating multilevel linear models as structural equation models." *Journal of Educational and Behavioral Statistics 28*:135—167.

Bauer, D. J. (2005, October). *Incongruence between the statistical theory and substantive application of growth mixture models in psychological research*. Cattell Award address presented at the annual meeting of the Society of Multivariate Experimental Psychology, Lake Tahoe, CA.

Bauer, D. J. (2007). "Observations on the use of growth mixture models in psychological research." *Multivariate Behavioral Research 42*:757—786.

Bauer, D. J. & Curran, P. J. (2002, June). *Estimating multilevel linear models as structural equation models*. Paper presented at the meeting of the Psychometric Society, Chapel Hill, NC.

Bauer, D. J. & Curran, P. J. (2003a). "Distributional assumptions of growth mixture models: Implications for overextraction of latent trajectory classes." *Psychological Methods 8*:338—363.

Bauer, D. J. & Curran, P. J. (2003b). "Over-extracting latent trajectory classes: Much ado about

nothing? Reply to Rindskopf(2003), Muthen(2003), and Cudeck and Henly(2003)." *Psychological Methods 8*:384—393.

Bauer, D. J. & Curran, P. J. (2004). "The integration of continuous and discrete latent variable models: Potential problems and promising opportunities." *Psychological Methods 9*:3—29.

Bell, R. Q. (1953). "Convergence: An accelerated longitudinal approach." *Child Development 24*: 145—152.

Bell, R. Q. (1954). "An experimental test of the accelerated longitudinal approach." *Child Development 25*:281—286.

Bentler, P. M. (1995). *EQS structural equations program manual*. Encino, CA: Multivariate Software.

Bentler, P. M. & Bonett, D. G. (1980). "Significance tests and goodness-of-fit in the analysis of covariance structures." *Psychological Bulletin 88*:588—606.

Biesanz, J. C., Deeb-Sossa, N., Papadakis, A. A., Bollen, K. A. & Curran, P. J. (2004). "The role of coding time in estimating and interpreting growth curve models." *Psychological Methods 9*:30—52.

Bijleveld, C. C. J. H. & van der Kamp, L. J. T. (1998). *Longitudinal data analysis: Designs, models, and methods*. Newbury Park, CA: Sage.

Blackson, T. C., Tarter, R. E., Loeber, R., Ammerman, R. T. & Windle, M. (1996). "The influence of paternal substance abuse and difficult temperament on sons' disengagement from family and deviant peers." *Journal of Youth and Adolescence 25*:389—411.

Blozis, S. A. (2004). "Structured latent curve models for the study of change in multivariate repeated measures." *Psychological Methods 9*:334—353.

Blozis, S. A. (2006). "A second-order structured latent curve model for longitudinal data." In K. van Montfort, H. Oud & A. Satorra(eds.), *Longitudinal models in the behavioral and related sciences*. Mahwah, NJ: Lawrence Erlbaum.

Blozis, S. A. (2007). "On fitting nonlinear latent curve models to multiple variables measured longitudinally." *Structural Equation Modeling 14*:179—201.

Bock, R. D. (1979). "Univariate and multivariate analysis of variance of time-structured data." In J. R. Nesselroade & P. B. Baltes(eds.), *Longitudinal research in the study of behavior and development*. New York: Academic Press.

Bollen, K. A. (1989). *Structural equations with latent variables*. New York: Wiley.

Bollen, K. A. & Curran, P. J. (2004). "Autoregressive latent trajectory(ALT) models: A synthesis of two traditions." *Sociological Methods & Research 32*:336—383.

Bollen, K. A. & Curran, P. J. (2006). *Latent curve models: A structural equation perspective*. Hoboken, NJ: Wiley.

Browne, M. W. (1993). "Structured latent curve models." In C. M. Cuadras & C. R. Rao(eds.), *Multivariate analysis: Future directions* 2:171—197. Amsterdam: Elsevier-North Holland.

Browne, M. W. & Cudeck, R. (1993). "Alternative ways of assessing model fit." In K. A. Bollen & J. S. Long(eds.), *Testing structural equation models*. Newbury Park, CA: Sage.

Browne, M. W. & du Toit, S. H. C. (1991). "Models for learning data." In L. M. Collins & J. L.

Horn(eds.), *Best methods for the analysis of change*. Washington, DC: American Psychological Association.

Browne, M. W., MacCallum, R. C., Kim, C., Andersen, B. L. & Glaser, R. (2002). "When fit indices and residuals are incompatible." *Psychological Methods 7*:403—421.

Bryk, A. S. & Raudenbush, S. W. (1987). Application of hierarchical linear models to assessing change. *Psychological Bulletin 101*:147—158.

Bryk, A. S. & Raudenbush, S. W. (1992). *Hierarchical linear Models: Applications and data analysis methods*. Newbury Park, CA: Sage.

Bryk, A. S., Raudenbush, S. W. & Congdon, R. T. (1996). *HLM: Hierarchical linear and non-linear modeling with the HLM/2L and HLM/3L programs*. Chicago: Scientific Software International.

Buist, K. L., Deković, M., Meeus, W. & van Aken, M. A. G. (2002). "Developmental patterns in adolescent attachment to mother, father and sibling." *Journal of Youth and Adolescence 31*: 167—176.

Byrne, B. M. & Crombie, G. (2003). "Modeling and testing change: An introduction to the latent growth curve model." *Understanding Statistics 2*:177—203.

Carrig, M. M., Wirth, R. J. & Curran, P. J. (2004). "A SAS macro for estimating and visualizing individual growth curves." *Structural Equation Modeling 11*:132—149.

Chan, D. (1998). "The conceptualization and analysis of change over time: An integrative approach incorporating longitudinal mean and covariance structures analysis(LMACS) and multiple indicator latent growth modeling(MLGM)." *Organizational Research Methods 1*:421—483.

Cheong, J., MacKinnon, D. P. & Khoo, S. T. (2003). "Investigation of mediational processes using parallel process latent growth curve modeling." *Structural Equation Modeling 10*:238—262.

Chou, C.-P., Bentler, P. M. & Pentz, M. A. (1998). "Comparisons of two statistical approaches to study growth curves: The multilevel model and the latent curve analysis." *Structural Equation Modeling 5*:247—266.

Cicchetti, D. & Rogosch, F. (1996). "Equifinality and multifinality in developmental psychopathology." *Development and Psychopathology 8*:597—600.

Coffman, D. L. & Millsap, R. E. (2006). "Evaluating latent growth curve models using individual fit statistics." *Structural Equation Modeling 13*:1—27.

Collins, L. M. (2006). "Analysis of longitudinal data: The integration of theoretical model, temporal design, and statistical model." *Annual Review of Psychology 57*:505—528.

Cronbach, L. J. & Webb, N. (1975). "Between class and within class effects in a reported aptitude× treatment interaction: A reanalysis of a study by G. L. Anderson." *Journal of Educational Psychology 67*:717—724.

Cudeck, R. & Klebe, K. J. (2002). "Multiphase mixed-effects models for repeated measures data." *Psychological Methods 7*:41—63.

Curran, P. J. (2000). "A latent curve framework for the study of developmental trajectories in adolescent substance use." In J. S. Rose, L. Chassin, C. C. Presson & S. J. Sherman(eds.), *Mult-*

*ivariate applications in substance use research*. Mahwah, NJ: Lawrence Erlbaum.

Curran, P. J. (2003). "Have multilevel models been structural equation models all along?" *Multivariate Behavioral Research 38*:529—569.

Curran, P. J., Bauer, D. J. &Willoughby, M. T. (2004). "Testing main effects and interactions in latent curve analysis." *Psychological Methods 9*:220—237.

Curran, P. J. & Bollen, K. A. (2001). "The best of both worlds: Combining autoregressive and latent curve models." In L. M. Collins & A. G. Sayer(eds.), *New methods for the analysis of change*. Washington, DC: American Psychological Association.

Curran, P. J., Harford, T. C. & Muthén, B. O. (1996). "The relation between heavy alcohol use and bar patronage: A latent growth model." *Journal of Studies on Alcohol 57*:410—418.

Curran, P. J. & Hussong, A. M. (2002). "Structural equation modeling of repeated measures data." In D. Moskowitz & S. Hershberger(eds.), *Modeling intraindividual variability with repeated measures data: Methods and applications*. New York: Lawrence Erlbaum.

Curran, P. J. & Hussong, A. M. (2003). "The use of latent trajectory models in psychopathology research." *Journal of Abnormal Psychology 112*:526—544.

Curran, P. J., Muthén, B. O. & Harford, T. C. (1998). "The influence of changes in marital status on developmental trajectories of alcohol use in young adults." *Journal of Studies on Alcohol 59*:647—658.

Curran, P. J., Stice, E. & Chassin, L. (1997). "The relation between adolescent alcohol use and peer alcohol use: A longitudinal random coefficients model." *Journal of Consulting and Clinical Psychology 65*:130—140.

Curran, P. J., West, S. G. & Finch, J. F. (1996). "The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis." *Psychological Methods 1*:16—29.

Curran, P. J. &Willoughby, M. T. (2003). "Implications of latent trajectory models for the study of developmental psychopathology." *Development and Psychopathology 15*:581—612.

de Leeuw, J. & Kreft, I. G. G. (2001). "Software for multilevel analysis." In A. H. Leyland & H. Goldstein(eds.), *Multilevelmodeling of health statistics*. Chichester, UK: Wiley.

Diggle, P. J., Liang, K.-Y. & Zeger, S. L. (1994). *Analysis of longitudinal data*. Oxford, UK: Clarendon Press.

du Toit, S. H. C. & Browne, M. W. (1992). AUFIT: Automated Fitting of Non-Standard Models [Computer software].

Duncan, S. C. & Duncan, T. E. (1994). "Modeling incomplete longitudinal substance use data using latent variable growth curve methodology." *Multivariate Behavioral Research 29*: 313—338.

Duncan, S. C. & Duncan, T. E. (1996). "A multivariate latent growth curve analysis of adolescent substance use." *Structural Equation Modeling 3*:323—347.

Duncan, S. C., Duncan, T. E. & Hops, H. (1996). "Analysis of longitudinal data within accelerated longitudinal design." *Psychological Methods 1*:236—248.

Duncan, T. E. & Duncan, S. C. (1995). "Modeling the processes of development via latent variable growth curve methodology." *Structural Equation Modeling 2*:187—213.

Duncan, T. E., Duncan, S. C. & Hops, H. (1993). "The effects of family cohesiveness and peer encouragement on the development of adolescent alcohol use: A cohort-sequential approach to the analysis of longitudinal data." *Journal of Studies on Alcohol 55*:588—599.

Duncan, T. E., Duncan, S. C. & Strycker, L. A. (2006). *An introduction to latent variable growth curve modeling: Concepts, issues, and applications* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum.

Duncan, T. E., Duncan, S., C., Strycker, L. A, Li, F. & Alpert, A. (1999). *An introduction tolatent variable growth curve modeling: Concepts, issues, and applications.* Mahwah, NJ: Lawrence Erlbaum.

Duncan, T. E., Tildesley, E., Duncan, S. C. & Hops, H. (1995). "The consistency of family and peer influences on the development of substance use in adolescence." *Addiction 90*: 1647—1660.

Enders, C. K. (2001). "A primer on maximum likelihood algorithms available for use with missing data." *Structural Equation Modeling 8*:128—141.

Enders, C. K. & Bandalos, D. L. (2001). "The relative performance of full information maximum likelihood estimation for missing data in structural equation models." *Structural Equation Modeling 8*:430—457.

FelTer, E., Hamagami, F. & McArdle, J. J. (2004). "Modeling latent growth curves with incomplete data using different types of structural equation modeling and multilevel software." *Structural Equation Modeling 11*:452—483.

Flora, D. B. & Chassin, L. (2005). "Changes in drug use during young adulthood: The effects of parent alcoholism and transition into marriage." *Psychology of Addictive Behaviors 19*:352—362.

Freud, A. (1958). "Adolescence." In R. S. Eissler, A. Freud, H. Hartmann & M. Kris(eds.), *The psychoanalytic study of the child* (Vol. 13:255—278). New York: International Universities Press.

George, R. (2003). "Growth in students' attitudes about the utility of science over the middle and high school years: Evidence from the Longitudinal Study of American Youth." *Journal of Science Education and Technology 12*:439—448.

Goldstein, H. (1995). *Multilevel statistical models* (2nd ed.). New York: Wiley.

Guttman, L. A. (1954). "A new approach to factor analysis: The radex." In P. E. Lazarsfeld(ed.), *Mathematical thinking in the social sciences.* New York: Columbia University Press.

Hamagami, E. (1997). "A review of the Mx computer program for structural equation modeling." *Structural Equation Modeling 4*:157—175.

Hancock. G. R. & Choi, J. (2006). "A vernacular for linear latent growth models." *Structural Equation Modeling 13*:352—377.

Hancock, G. R., Kuo, W. -L. & Lawrence. F. R. (2001). "An illustration of second-order latent growth models." *Structural Equation Modeling 8*:470—489.

Hancock, G. R. & Lawrence, E. R. (2006). "Using latent growth models to evaluate longitudinal change." In G. R. Hancock & R. O. Mueller(eds.), *Structural equation modeling: A second*

course. Greenwich, CT: Information Age.

Hertzog, C., Lindenberger, U., Ghisletta, P. & von Oertzen, T. (2006). "On the power of multivariate latent growth curve models to detect correlated change." *Psychological Methods* 11: 244—252.

Hofmann, D. A. (1997). "An overview of the logic and rationale of hierarchical linear models." *Journal of Management* 23:723—744.

Hox, J. (1998). "Multilevel modeling: When and why." In I. Balderjahn, R. Mathar & M. Schader (eds.). *Classification, data analysis, and data highways*. Berlin: Springer.

Hox, J. (2000). "Multilevel analyses of grouped and longitudinal data." In T. D. Little, K. U. Schnabel & J. Baumert(eds.), *Modeling longitudinal and multilevel data: Practical Issues, applied approaches and specific examples*. Mahwah, NJ: Lawrence Erlbaum.

Hox, J. (2002). *Multilevel analysis: Techniques and applications*. Mahwah, NJ: Lawrence Erlbaum.

Hussong, A. M., Hicks, R. E., Levy, S. A. & Curran, P. J. (2001). "Specifying the relations between affect and heavy alcohol use among young adults." *Journal of Abnormal Psychology 110*:449—461.

Jones. B. L., Nagin. D. S. & Roeder, K. (2001). "A SAS procedure based on mixture models for estimating developmental trajectories." *Sociological Methods & Research* 29:374—393.

Jöreskog, K. G. (1967). "Some contributions to maximum likelihood factor analysis." *Psysmetrika* 32:443—482.

Jöreskog, K. G. (2002). *Structural Equation Modeling with ordinal variables using LISREL*. Retrieved January 17, 2008, from http://www.ssicentral.com.

Joreskog, K. G. & Sorbom, D. (1996). *LISREL 8 user's reference guide*. Chicago: Scientific Software International.

Khoo, S.-T. & Muthén, B. (2000). "Longitudinal data on families: Growth modeling alternatives." In J. S. Rose, L. Chassin, C. C. Presson & S. J. Sherman(eds.), *Multivariate applications in substance use research*. Mahwah, NJ: Lawrence Erlbaum.

Kiecolt-Glaser, J. K., Glaser, R., Cacioppo, J. T., MacCallum, R. C., Snydersmith, M., Kim, C., et al. (1997). "Marital conflict in older adults: Endocrinological and immunological correlates." *Psychosomatic Medicine 59*:350—351.

Klein, A. G. & Muthén, B. O. (2006). "Modeling heterogeneity of latent growth depending on initial status." *Journal of Educational and Behavioral Statistics 31*:357—375.

Kline, R. B. (2004). *Principles and practice of structural Equation Modeling* (2nd ed.). New York: Guilford Press.

Kreft, I. G. G. & de Leeuw, J. (1998). *Introducing multilevel modeling*. London: Sage.

Lawrence, E. R. & Hancock, G. R. (1998). "Assessing change over time using latent growth modeling." *Measurement and Evaluation in Counseling and Development 30*:211—224.

Lee, S.-Y. (2007). *Structural Equation Modeling: A Bayesian approach*. Hoboken, NJ: Wiley.

Lei, M. & Lomax, R. G. (2005). "The effect of varying degrees of nonnormality in structural equation modeling." *Structural Equation Modeling* 12:1—27.

Li, F., Duncan, T. E. & Acock. A. (2000). "Modeling interaction effects in latent growth curve models." *Structural Equation Modeling* 7:497—533.

Li, F., Duncan, T. E., Duncan, S. C. & Acock, A. (2001). "Latent growth modeling of longitudinal data: A finite growth mixture modeling approach." *Structural Equation Modeling* 8:493—530.

Li, F., Duncan, T. E., Harmer, P., Acock, A. & Stoolmiller, M. (1998). "Analyzing measurement models of latent variables through multilevel confirmatory factor analysis and hierarchical linear modeling approaches." *Structural Equation Modeling* 5:294—306.

Liu, H. & Powers, D. A. (2007). "Growth curve models for zero-inflated count data: An application to smoking behavior." *Structural Equation Modeling* 14:247—279.

Longford, N. T. (1993). *Random coefficient models*. Oxford, UK: Clarendon Press.

Luke, D. A. (2004). *Multilevel Modeling*. Thousand Oaks, CA: Sage.

MacCallum, R. C., Browne, M. W. & Sugawara, H. M. (1996). "Power analysis and determination of sample size for covariance structure modeling." *Psychological Methods* 1:130—149.

MacCallum, R. C. & Kim. C. (2000). "Modeling multivariate change." In T. D. Little, K. U. Schnabel & J. Baumert (eds.), *Modeling longitudinal and multilevel data: Practical issue, applied approaches and specific examples* (pp. 51—68). Mahwah, NJ: Lawrence Erlbaum.

MacCallum, R. C., Kim, C., Malarkey, W. B. & Kiecolt-Glaser, J. K. (1997). "Studying multivariate change using multilevel models and latent curve models." *Multivariate Behavioral Research* 32:215—253.

MacCallum, R. C., Roznowski, M. & Necowitz, L. B. (1992). "Model modification in covariance structure analysis: The problem of capitalization on chance." *Psychological Bulletin* 111: 490—504.

Marini, M. M., Olsen, A. R. & Rubin, D. R(1979). "Maximum likelihood estimation in panel studies with missing data." In K. E. Schuessler (ed.), *Sociological methodology 1980* (pp. 314—357). San Francisco: Jossey-Bass.

Maruyama, G. M. (1997). *Basics of structural equation modeling*. Thousand Oaks, CA: Sage.

McArdle, J. J. (1988). "Dynamic but structural equation modeling of repeated measures data." In R. B. Cattell & J. Nesselroade (eds.), *Handbook of multivariate experimental psychology* (2nd ed.). New York: Plenum.

McArdle. J. J. (1989). "A structural modeling experiment with multiple growth functions." In R. Kanfer, P. L. Ackerman & R. Cudeck (eds.), *Abilities, motivation, and methodology: The Minneapolis symposium on learning and individual differences* (pp. 71—117). Hillsdale, NJ: Lawrence, Erlbaum.

McArdle, J. J. (2001). "A latent difference score approach to longitudinal dynamic structural analyses." In R. Cudeck, S. du Toit & D. Sörbom (eds.), *Structural equation modeling: Present and future—A festschrift in honor of Karl Jöreskog*. Lincolnwood, IL: Scientific Software International.

McArdle, J. J. & Anderson, E. (1990). "Latent variable growth models for research on aging." In J. E. Birren & K. W. Schaie (eds.), *Handbook of the psychology of aging* (3rd ed.). San Die-

go, CA: Academic Press.

McArdle, J. J. & Bell, R. Q. (2000). "An introduction to latent growth models for developmental data analysis." In T. D. Little, K. U. Schnabel & J. Baumert (eds.), *Modeling longitudinal and multilevel data: Practical issues, applied approaches and specific examples*. Mahwah, NJ: Lawrence Erlbaum.

McArdle, J. J. & Epstein, D. (1987). "Latent growth curves within developmental structural equation models." *Child Development 58*:110—133.

McArdle, J. J. & Hamagami, F. (1991). "Modeling incomplete longitudinal data using latent growth structural equation models." In L. Collins & J. L. Horn (eds.), *Best methods for the analysis of change*. Washington, DC: American Psychological Association.

McArdle, J. J. & Hamagami, F. (1992). "Modeling incomplete longitudinal and cross-sectional data using latent growth structural models." *Experimental Aging Research 18*:145—166.

McArdle, J. J. & Hamagami, F. (2001). "Latent difference score structural models for linear dynamic analyses with incomplete longitudinal data." In L. M. Collins & A. G. Sayer (eds.), *New methods for the analysis of change*. Washington, DC: American Psychological Association.

Mehta, P. D. & Neale, M. C. (2005). "People are variables too: Multilevel structural equation modeling." *Psychological Methods 10*:259—284.

Mehta, P. D., Neale, M. C. & Flay, B. R. (2004). "Squeezing interval change from ordinal panel data: Latent growth curves with ordinal outcomes." *Psychological Methods* 9:301—333.

Mehta, P. D. & West, S. G. (2000). "Putting the individual back into individual growth curves." *Psychological Methods* 5:23—43.

Meredith, W. & Horn, J. (2001). "The role of factorial invariance in modeling growth and change." In L. M. Collins & A. G. Sayer (eds.), *New methods for the analysis of change*. Washington, DC: American Psychological Association.

Meredith, W. & Tisak, J. (1990). "Latent curve analysis." *Psychometrika 55*:107—122.

Miller, B. C., Benson, B. & Galbraith, K. A. (2001). "Family relationship and adolescent pregnancy risk: A research synthesis." *Developmental Review 21*:1—38.

Miyazaki, Y. & Raudenbush, S. W. (2000). "Tests for linkage of multiple cohorts in an accelerated longitudinal design." *Psychological Methods 5*:44—63.

Muthén, B. (1993). "Latent variable modeling of growth with missing data and multilevel data." In C. M. Cuadras & C. R. Rao (eds.), *Multivariate analysis: Future directions 2*: 199—210. Amsterdam: North Holland.

Muthén, B. (1997). "Latent variable modeling of longitudinal and multilevel data." In A. Raftery (ed.), *Sociological methodology*. Boston: Blackwell.

Muthén, B. (2000). "Methodological issues in random coefficient growth modeling using a latent variable framework: Applications to the development of heavy drinking ages 18—37." In J. S. Rose, L. Chassin, C. C. Presson & S. J. Sherman (eds.), *Multivariate applications in substance use research*. Mahwah, NJ: Lawrence Erlbaum.

Muthén, B. (2001). "Latent variable mixture modeling." In G. A. Marcoulides & R. E. Schumacker (eds.), *New developments and techniques in structural Equation Modeling*. Mahwah, NJ:

Lawrence Erlbaum.

Muthén, B. & Asparouhov, T. (2002). "Latent variable analysis with categorical outcomes: Multiple-group and growth modeling in Mplus." *Mplus Web Note: No. 4*. Retrieved January 17, 2008, from http://www.statmodel.com/download/webnotes/CatMGLong.pdf.

Muthén, B., Kaplan, D. & Hollis, M. (1987). "On structural equation modeling with data that are not missing completely at random." *Psychometrika 52*:431—462.

Muthén, B. O. & Curran, P. J. (1997). "General longitudinal modeling of individual differences in experimental designs: A latent variable framework for analysis and power estimation." *Psychological Methods 2*:371—402.

Muthén, L. K. & Muthén, B. O. (1998—2006). *Mplus user's guide*. Los Angeles: Muthén & Muthén.

Nagin, D. (1999). "Analyzing developmental trajectories: A semi-parametric, group-based approach." *Psychological Methods 4*:139—177.

Nagin, D. & Tremblay, R. E. (2001). "Analyzing developmental trajectories of distinct but related behaviors: A group-based method." *Psychological Methods 6*:18—34.

Neale, M. C. (2000). "Individual fit, heterogeneity, and missing data in multi group structural equation modeling." In T. D. Little, K. U. Schnabel & J. Baumert(eds.), *Modeling Longitudinal and multilevel data: Practical issues, applied approaches and specific examples*. Mahwah, NJ: Lawrence Erlbaum.

Neale, M. C., Boker, S. M., Xie, G. & Maes, H. H. (2003). *Mx: Statistical modeling* (6th ed.). Richmond, VA: Department of Psychiatry, Virginia Commonwealth University.

Neale, M. C. & Miller, M. B. (1997). "The use of likelihood-based confidence intervals in genetic models." *Behavior Genetics 27*:113—120.

Nesselroade, J. R. & Baltes, P. B. (1979). *Longitudinal research in the study of behavior and development*. New York: Academic Press.

NICHD Early Child Care Research Network. (2006). Child-care effect sizes for the NICHD study of early child care and youth development. *American Psychologist 61*:99—116.

Pianta, R. C. (1993). *The Student-Teacher Relationship Scale*. Charlottesville: University of Virginia Press.

Potthoff, R. F. &Roy, S. N. (1964). "A generalized multivariate analysis of variance model useful especially for growth curve problems." *Biometrika 51*:313—326.

Preacher, K. J., Curran, P. J. & Bauer, D. J. (2006). "Computational tools for probing interaction effects in multiple linear regression, multilevel modeling, and latent curve analysis." *Journal of Educational and Behavioral Statistics 31*:437—448.

Rao, C. R. (1958). "Some statistical models for comparison of growth curves." *Biometrics 14*: 1—17.

Rasbash, J., Browne, W., Goldstein, H., Yang, M., Plewis, I., Healy, M., et al. (1999). *A user's guide to MLwiN*. London: Multilevel Models Project.

Raudenbush, S. W. (2001). "Toward a coherent framework for comparing trajectories of individual change." In L. M. Collins & A. G. Sayers(eds.), *New methods for the analysis of change*.

Washington, DC: American Psychological Association.

Raudenbush, S. W. , Brennan, R. T. & Barnett, R. C. (1995). "A multivariate hierarchical model for studying psychological change within married couples." *Journal of Family Psychology 9*: 161—176.

Raudenbush, S. W. & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage.

Raudenbush, S. W. & Chan, W. (1992). "Growth curve analysis in accelerated longitudinal designs." *Journal of Research in Crime and Delinquency 29*:387—411.

Raudenbush, S. W. & Chan, W. -S. (1993). "Application of a hierarchical linear model to the study of adolescent deviance in an overlapping cohort design." *Journal of Consulting and Clinical Psychology 61*:941—951.

Raykov, T. & Marcoulides, G. A. (2000). *A first course in structural Equation modeling*. Hillsdale, NJ: Lawrence Erlbaum.

Rodebaugh, T. L. , Curran, P. J. & Chambless, D. L. (2002). "Expectancy of panic in the maintenance of daily anxiety in panic disorder with agoraphobia: A longitudinal test of competing models." *Behavior Therapy 33*:315—336.

Rogosa, D. R. & Willett, J. B. (1985). "Understanding correlates of change by modeling individual differences in growth." *Psychometrika 50*:203—228.

Rovine, M. J. & Molenaar, P. C. M. (2000). "A structural modeling approach to a multilevel random coefficients model." *Multivariate Behavioral Research 35*:51—88.

Rubin, D. B. (1976). "Inference and missing data." *Biometrika 63*:581—592.

Sayer, A. G. & Cumsille, P. E. (2001). "Second-order latent growth models." In L. M. Collins & A. G. Sayer (eds.), *New methods for the analysis of change*. Washington, DC: American Psychological Association.

Sayer, A. G. & Willett, J. B. (1998). "A cross-domain model for growth in adolescent expectancies." *Multivariate Behavioral Research 33*:509—543.

Schaie, K. W. (1965). "A general model for the study of developmental problems." *Psychological Bulletin 64*:92—107.

Schaie, K. W. (1986). "Beyond calendar definitions of age, time, and cohort: The general developmental model revisited." *Developmental Review 6*:252—277.

Scheines, R. , Hoijtink, H. & Boomsma, A. (1999). "Bayesian estimation and testing of structural equation models." *Psychometrika 64*:37—52.

Simons-Morton, B. G. , Chen, R. , Abroms, R. & Haynie, D. L. (2004). "Latent growth curve analyses of peer and parent influences on smoking stage progression among early adolescents." *Health Psychology 23*:612—621.

Singer, J. D. & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. New York: Oxford University Press.

Smetana, J. G. (1988). "Concepts of self and social convention: Adolescents' and parents' reasoning about hypothetical and actual family conflicts." In M. R. Gunnar & W. A. Collins (eds.), *The Minnesota Symposia on Child Psychology: Vol. 21. Development during the transition to ado-*

lescence. Hillsdale, NJ: Lawrence Erlbaum.

Snijders, T. & Bosker, R. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. London: Sage.

Stack, S. & Eshleman, J. R. (1998). "Marital status and happiness: A 17-nation study." *Journal of Marriage and the Family 60*:527—536.

Steiger, J. H. & Lind, J. M. (1980, June). *Statistically based tests for the number of common factors*. Paper presented at the annual meeting of the Psychometric Society, Iowa City, IA.

Steinberg, L. (1989). "Pubertal maturation and parent-adolescent distance: An evolutionary perspective." In G. Adams, R. Montemayor & T. Gullotta(eds.), *Biology of adolescent behaviorand development*. Newbury Park, CA: Sage.

Sterba, S., Prinstein, M. J. & Cox, M. J. (2007). "Trajectories of internalizing problems across childhood: Heterogeneity, external validity, and gender differences." *Development and Psychopathology 19*:345—366.

Stoel, R. D. (2003). *Issues in growth curve modeling*. Unpublished doctoral dissertation, University of Amsterdam, The Netherlands.

Stoel, R. D. & van den Wittenboer, G. (2003). "Time dependence of growth parameters in latent growth curve models with time invariant covariates." *Methods of Psychological Research Online 8*:21—41.

Stoel, R. D., van den Wittenboer, G. & Hox, J. (2004). "Including time-invariant covariates in the latent growth curve model." *Structural Equation Modeling 11*:155—167.

Stoolmiller, M. (1995). "Using latent growth curve models to study developmental processes." In J. M. Gottman(ed.), *The analysis of change*. Mahwah, NJ: Lawrence Erlbaum.

Tisak, J. & Meredith, W. (1989). "Exploratory longitudinal factor analysis in multiple populations." *Psychometrika 54*:261—281.

Tisak, J. & Meredith, W. (1990). "Descriptive and associative developmental models." In A. von Eye (ed.), *Statistical methods in longitudinal research*(Vol. 2, 387—406). Boston: Academic Press.

Tisak, J. & Tisak, M. S. (1996). "Longitudinal models of reliability and validity: A latent curve approach." *Applied Psychological Measurement 20*:275—288.

Tisak, J. & Tisak, M. S. (2000). "Permanency and ephemerality of psychological measures with application to organizational commitment." *Psychological Methods 5*:175—198.

Tonry, M., Ohlin, L. E. & Farrington, D. P. (1991). *Human development and criminal behavior: New ways of advancing knowledge*. New York: Springer-Verlag.

Tucker, L. R. (1958). "Determination of parameters of a functional relation by factor analysis." *Psychometrika 23*:19—23.

Tucker, L. R. (1966). "Learning theory and multivariate experiment: Illustration by determination of parameters of generalized learning curves." In R. B. Cattell(ed.), *Handbook of multivariate experimental psychology*. Chicago: Rand McNally.

Tucker, L. R. & Lewis, C. (1973). "A reliability coefficient for maximum likelihood factor analysis." *Psychometrika 38*:1—10.

VanLaningham, J., Johnson, D. R. & Amato, P. (2001). "Marital happiness, marital duration, and the

U-shaped curve: Evidence from a five-wave panel study." *Social Forces 78*:1313—1341.

von Eye, A. & Bergman, L. R. (2003). "Research strategies in developmental psychopathology: Dimensional identity and the person-oriented approach." *Development and Psychopathology 15*:553—580.

Wang, J. (2004). "Significance testing for outcome changes via latent growth model." *Structural Equation Modeling 11*:375—400.

Wang, M. & Bodner, T. E. (2007). "Growth mixture modeling: Identifying and predicting unobserved subpopulations with longitudinal data." *Organizational Research Methods 10*: 635—656.

Wendorf, C. A. (2002). "Comparisons of structural equation modeling and hierarchical linear modeling approaches to couples' data." *Structural Equation Modeling 9*:126—140.

Widaman, K. F. & Thompson, J. S. (2003). "On specifying the null model for incremental fit indices in structural equation modeling." *Psychological Methods 8*:16—37.

Willett, J. B. (1989). "Questions and answers in the measurement of change." In E. Z. Rothkopf (ed.), *Review of research in education* (Vol. 15). Washington, DC: American Education Research Association.

Willett, J. B. & Sayer, A. G. (1994). "Using covariance structure analysis to detect correlates and predictors of individual change over time." *Psychological Bulletin 116*:363—381.

Willett, J. B. & Sayer, A. G. (1995). "Cross-domain analyses of change over time: Combining growth modeling and covariance structure analysis." In G. A. Marcoulides & R. E. Schumacker (eds.), *Advanced structural equation modeling: Issues and techniques*. Mahwah, NJ: Lawrence Erlbaum.

Willett, J. B., Singer, J. D. & Martin, N. C. (1998). "The design and analysis of longitudinal studies of development and psychopathology in context: Statistical models and methodological recommendations." *Development and Psychopathology 10*:395—426.

Wills, T. A., Cleary, S. D., Filer, M., Shinar, O., Mariani, J. & Spera, K. (2001). "Temperament related to early-onset substance use: Test of a developmental model." *Prevention Science 2*:145—163.

Wothke, W. (2000). "Longitudinal and multigroup modeling with missing data." In T. D. Little, K. U. Schnabel & J. Baumert (eds.), *Modeling longitudinal and multilevel data: Practical issues, applied approaches and specific examples*. Mahwah, NJ: Lawrence Erlbaum.

Zhang, Z., Hamagami, F., Wang, L., Nesselroade, J. R. & Grimm, K. J. (2007). "Bayesian analysis of longitudinal data using growth curve models." *International Journal of Behavioral Development 31*:374—383.